COLLEGE OF ENGINEERING
COMPUTER SCIENCE & ENGINEERING
UNIVERSITY OF MICHIGAN

GEMS LAB

Lawrence Livermore National Laboratory

Network Embedding for Role Discovery: Concepts, Tools, and Applications

Mark Heimann, Junchen Jin, Danai Koutra

SIAM International Conference on Data Mining– April 30, 2022

About the tutorial organizers

Mark Heimann
LLNL

Junchen Jin
Northwestern University

Danai Koutra
University of Michigan,
Amazon

Tutorial Goal

This tutorial aims to provide an (incomplete) overview of the major structural or role-based node embedding methods, and connect them to role equivalence research in mathematical sociology.

Expectations:

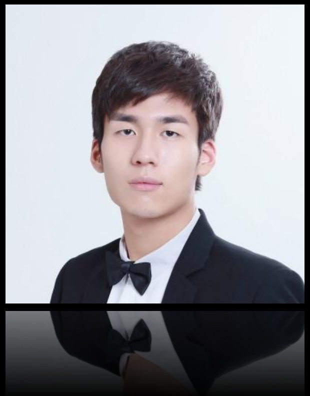
- Familiarity with graphs and representation learning.
- We'll provide a high-level introduction and relevant definitions.

Tutorial Outline:
Network Embedding for Role Discovery

- Part I: Lecture
  - Introduction
  - Structural roles in
    - network science
    - mathematical sociology
  - Structural or role-based embedding methods
  - Mining structural roles within a network
  - Mining structural roles across networks
- Part II: Demo
  - Hands-on demo

Tutorial Outline:
Network Embedding for Role Discovery

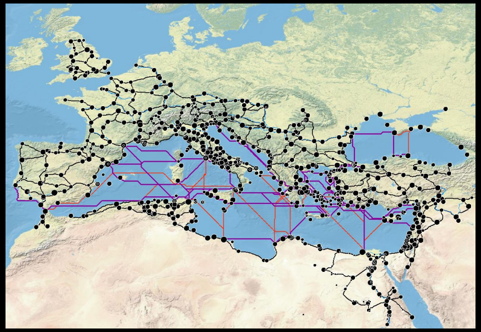
- Part I: Lecture
  - Introduction
  - Structural roles in
    - network science
    - mathematical sociology
  - Structural or role-based embedding methods
  - Mining structural roles within a network
  - Mining structural roles across networks
- Part II: Demo
  - Hands-on demo

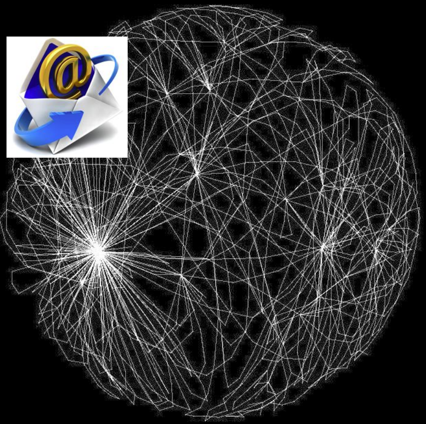


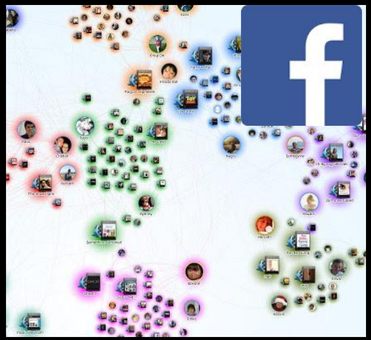
Tutorial Outline:
Network Embedding for Role Discovery

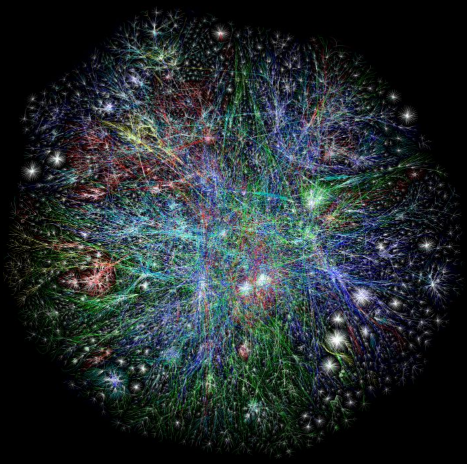
- Part I: Lecture
  - **Introduction**
  - Structural roles in
    - network science
    - mathematical sociology
  - Structural or role-based embedding methods
  - Mining structural roles within a network
  - Mining structural roles across networks
- Part II: Demo
  - Hands-on demo

Networks Are Everywhere





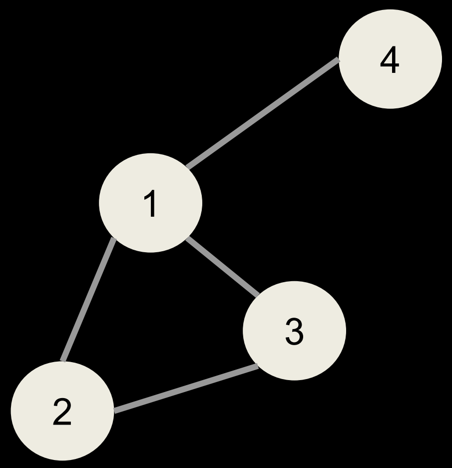




Networks: The Basics

Graph (or network)

$$G = (V, E \subseteq V^2)$$



Adjacency matrix **A**

$$\begin{array}{cc} & \text{Nodes} \\ & \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \text{Nodes} \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array} & \left[\begin{array}{cccc} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{array}\right] \end{array}$$

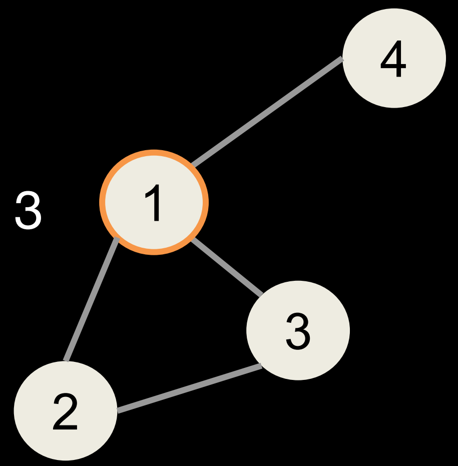
edge between nodes 1 and 2 (and 2,1 since graph is undirected)

Node Degree

Number of neighbors or connections

- Most basic, simple to compute
- Highly descriptive of structural role

degree
$deg(v_1) = 3$



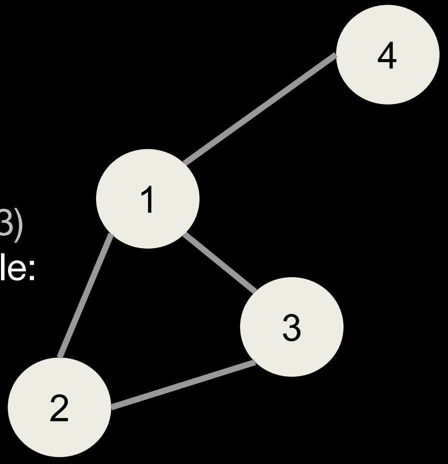
(Local) Clustering Coefficient

Proportion of triangles in neighborhood

- Tells how clique-like the node's neighborhood is

$CC(v_1) = ⅓$

1 triangle (1-2-3)
out of 3 possible:
(1-2-3, 1-2-4,
1-3-4)



\# of triangles in u's neighborhood

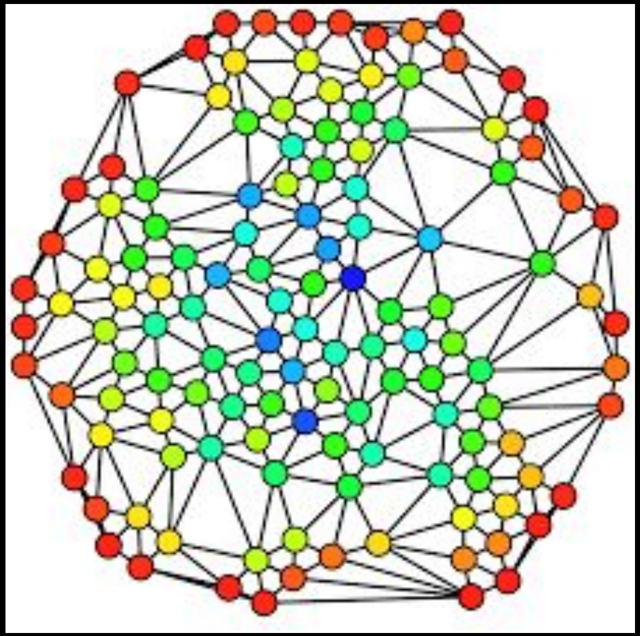
$$CC(u) = \frac{|\{i,j \in \mathcal{N}(u) : \mathbf{A_{ij}} > 0\}|}{deg(u)\big(deg(u)-1\big)/2}$$

\# of possible triangles

Betweenness Centrality

Portion of shortest paths going through node

- Measures the “monitoring” role of the node
- High centrality means the node is essential for passing information through the network
- More global, also more expensive to compute



$$BC(u) = \sum_{i,j:i,j \neq u} \frac{\#SP(i,j;u)}{\#SP(i,j)}$$

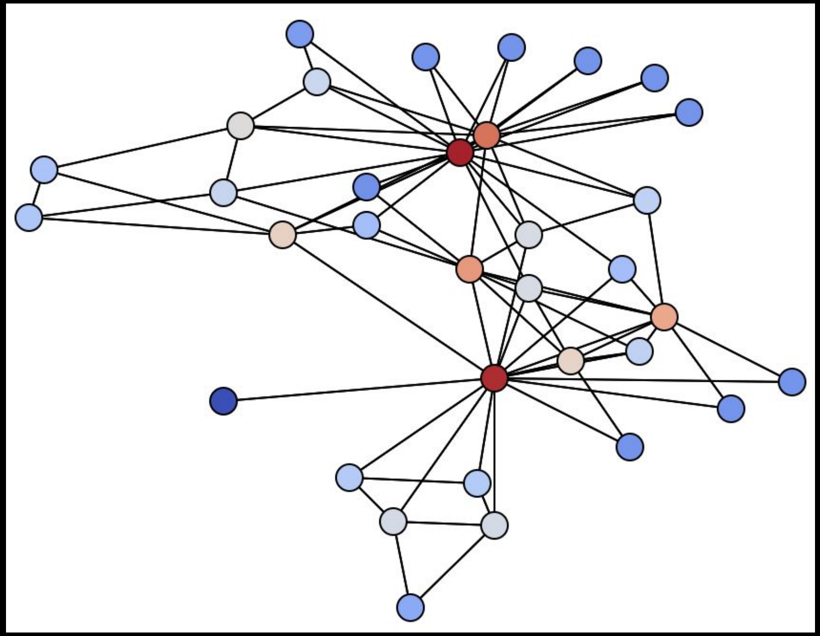
PageRank

Google

Iterative computation of influence/importance scores

- More influential nodes are linked to by other influential nodes
- Links count more the fewer of them a node sends out

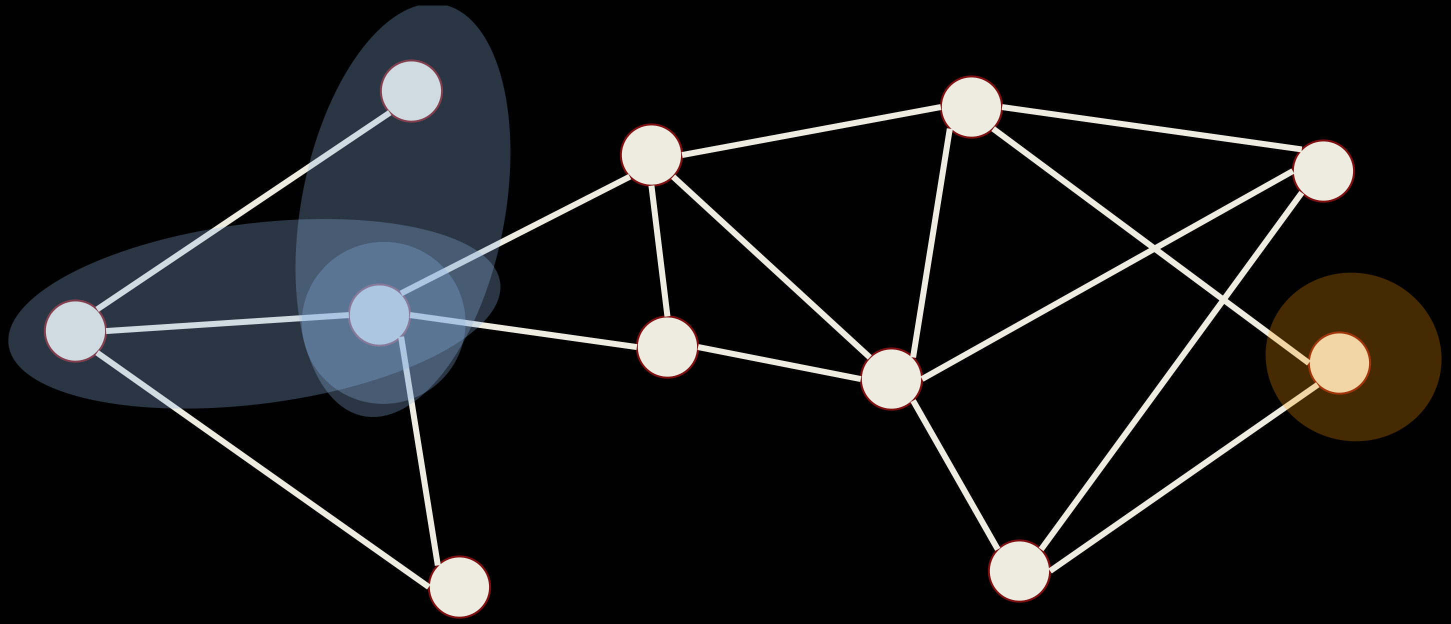
$$\boxed{PR_t = \underbrace{\delta(\mathbf{D}^{-1}\mathbf{A})PR_{t-1}}_{\text{with prob } \delta \text{ follow a link at random}} + \underbrace{\frac{1-\delta}{N}\mathbf{1}}_{\text{with prob } 1-\delta \text{ teleport to a random node}}}$$



Tutorial Outline

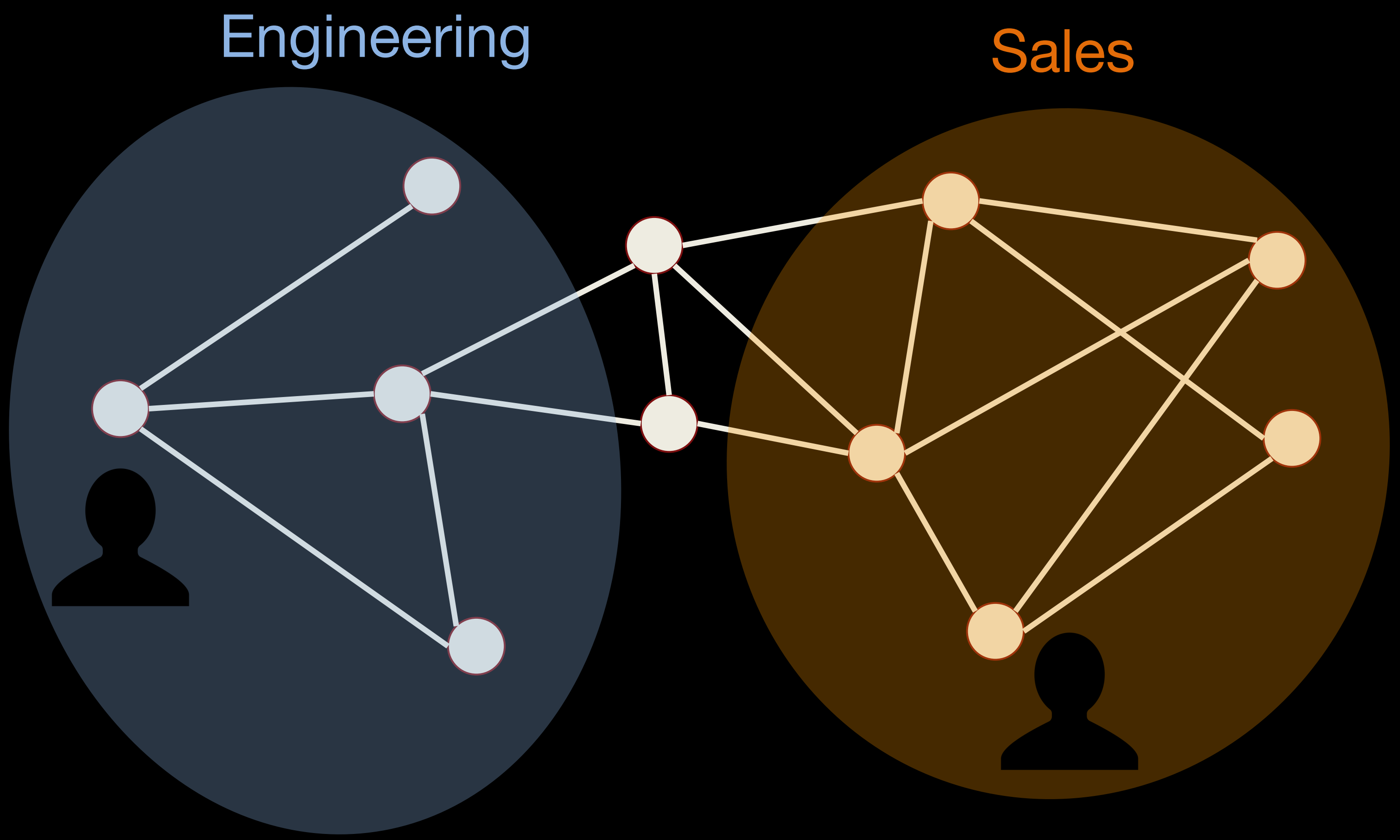
- Part I: Lecture
  - Introduction
  - **Structural roles in**
    - **network science**
    - mathematical sociology
  - Structural or role-based embedding methods
  - Mining structural roles within a network
  - Mining structural roles across networks
- Part II: Demo
  - Hands-on demo

Which Parts of a Network are Similar?



What Can We Learn from Network Similarity?

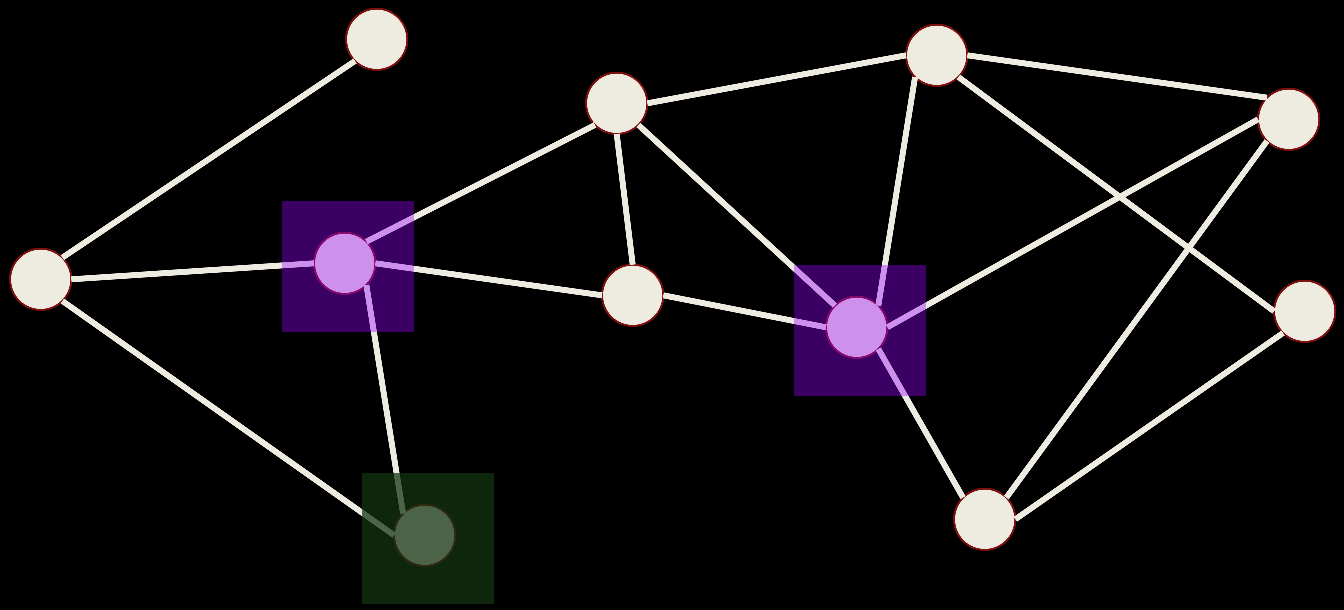
POV: this network models a company's internal communication



Which Parts of a Network are Similar?

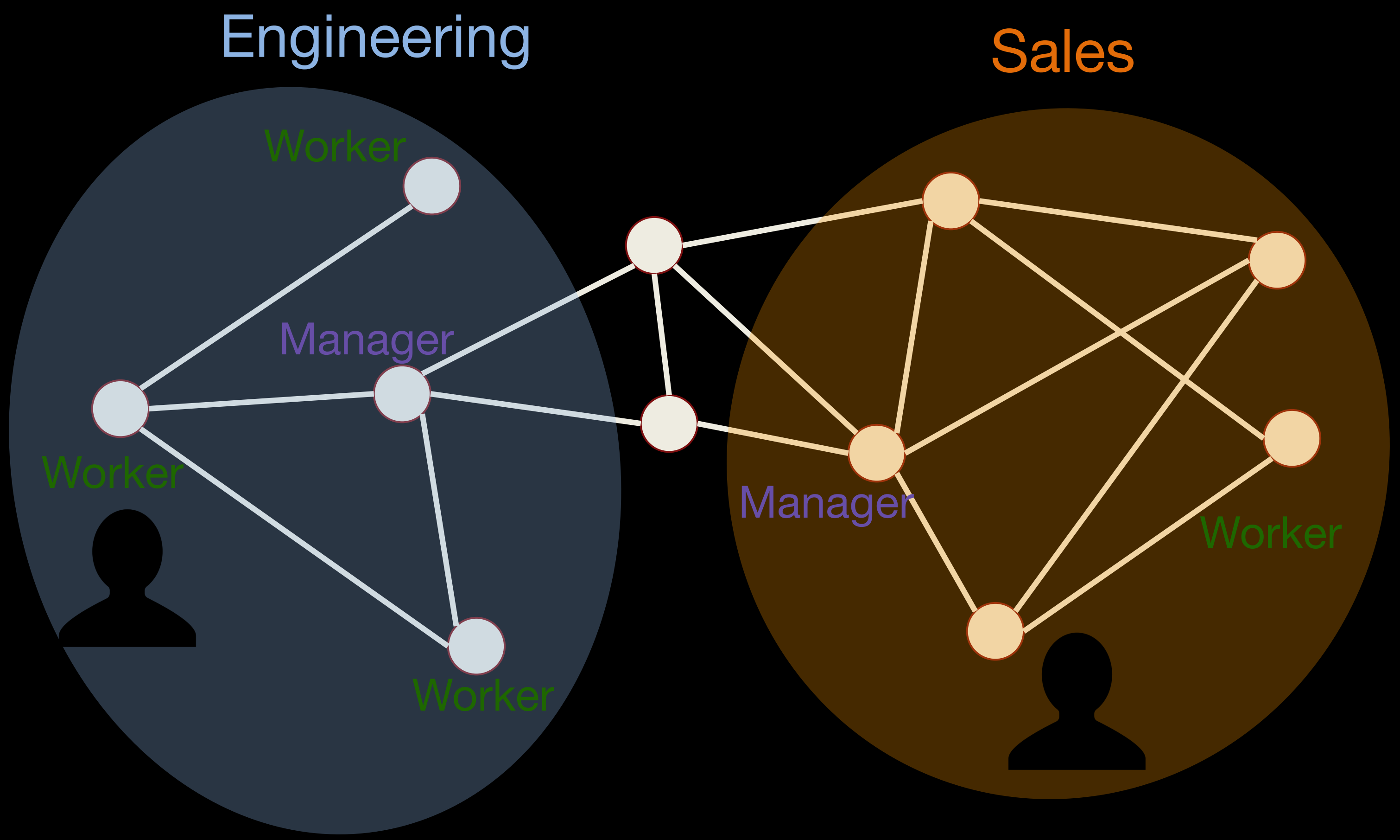
Similar *patterns* of connectivity

- Which may not be true of nodes that are connected!

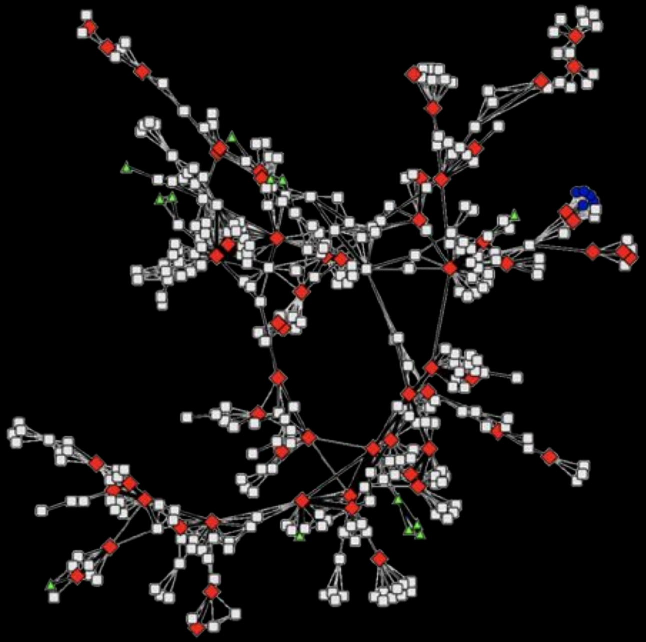


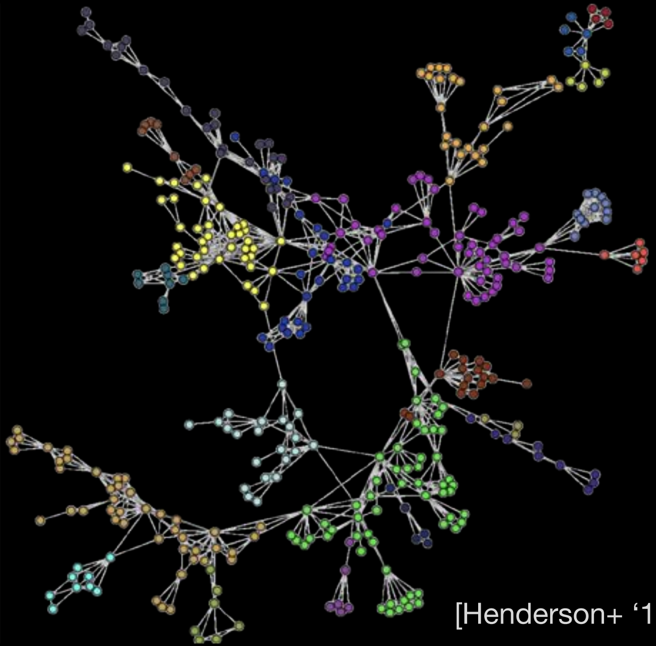
What Can We Learn from Network Similarity?

POV: this network models a company’s internal communication



Structural Similarity vs. Proximity





[Henderson+ ‘12]

Structural Similarity

- Find similarity between nodes all over the network with similar roles
- Useful for role-based classification
- Can be compared across networks

[Jin+ ‘21] [Rossi+ ‘21]

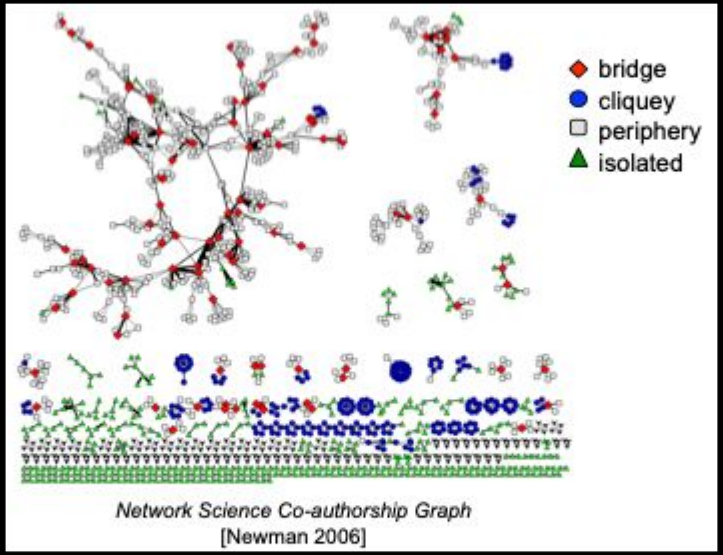
Proximity

- Find similarity only between nodes in the same part of the network
- Useful for link prediction, classification when labels exhibit homophily

[Grover+ ‘16; Perozzi+ ‘14]

What are roles?

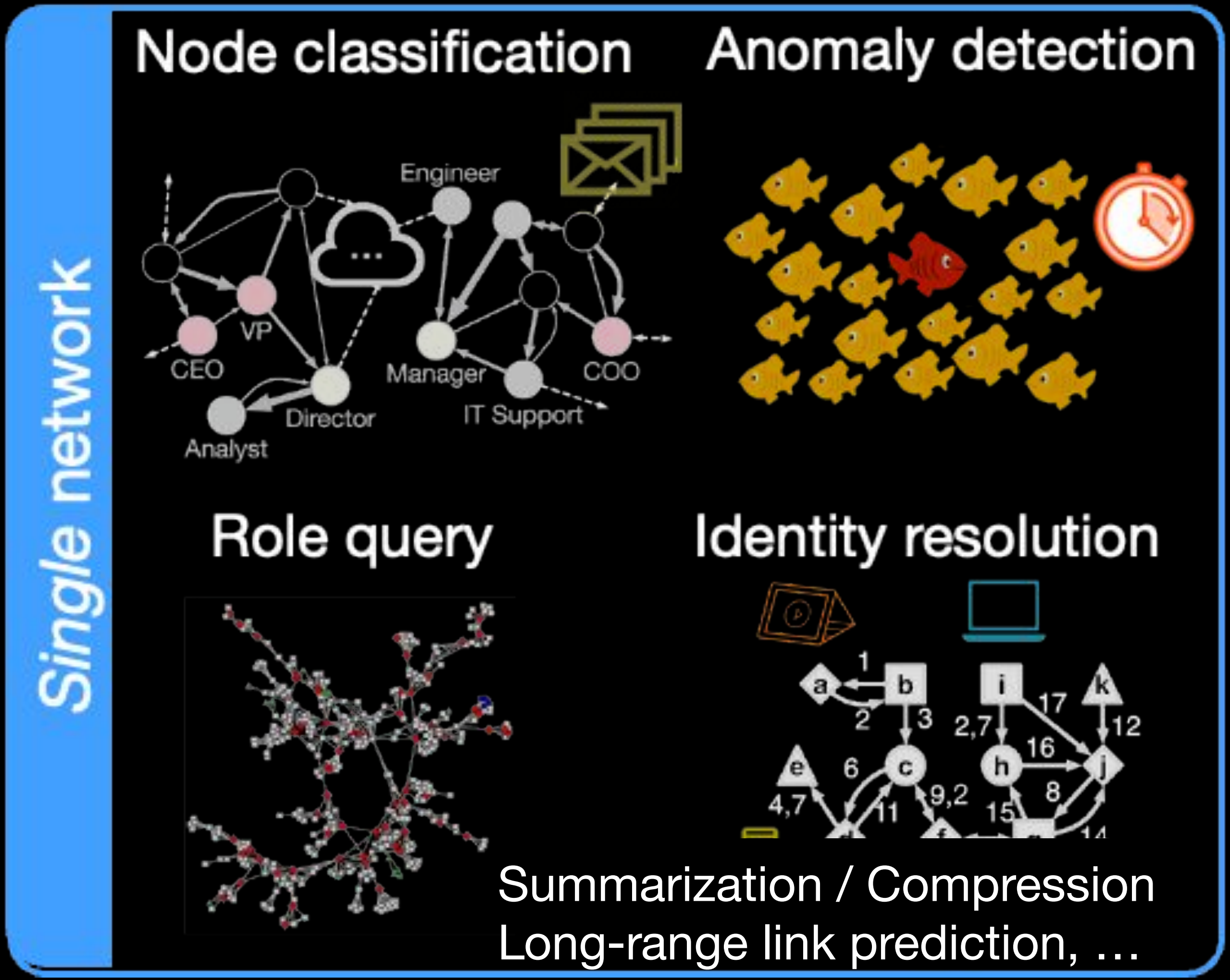
- The ways in which nodes / entities / actors relate to each other
- "The behavior expected of a node occupying a specific position" [Homans '67]
  - ✧ e.g., centers of stars
  - ✧ members of cliques
  - ✧ peripheral nodes
- Equivalence class: collection of nodes with the same role

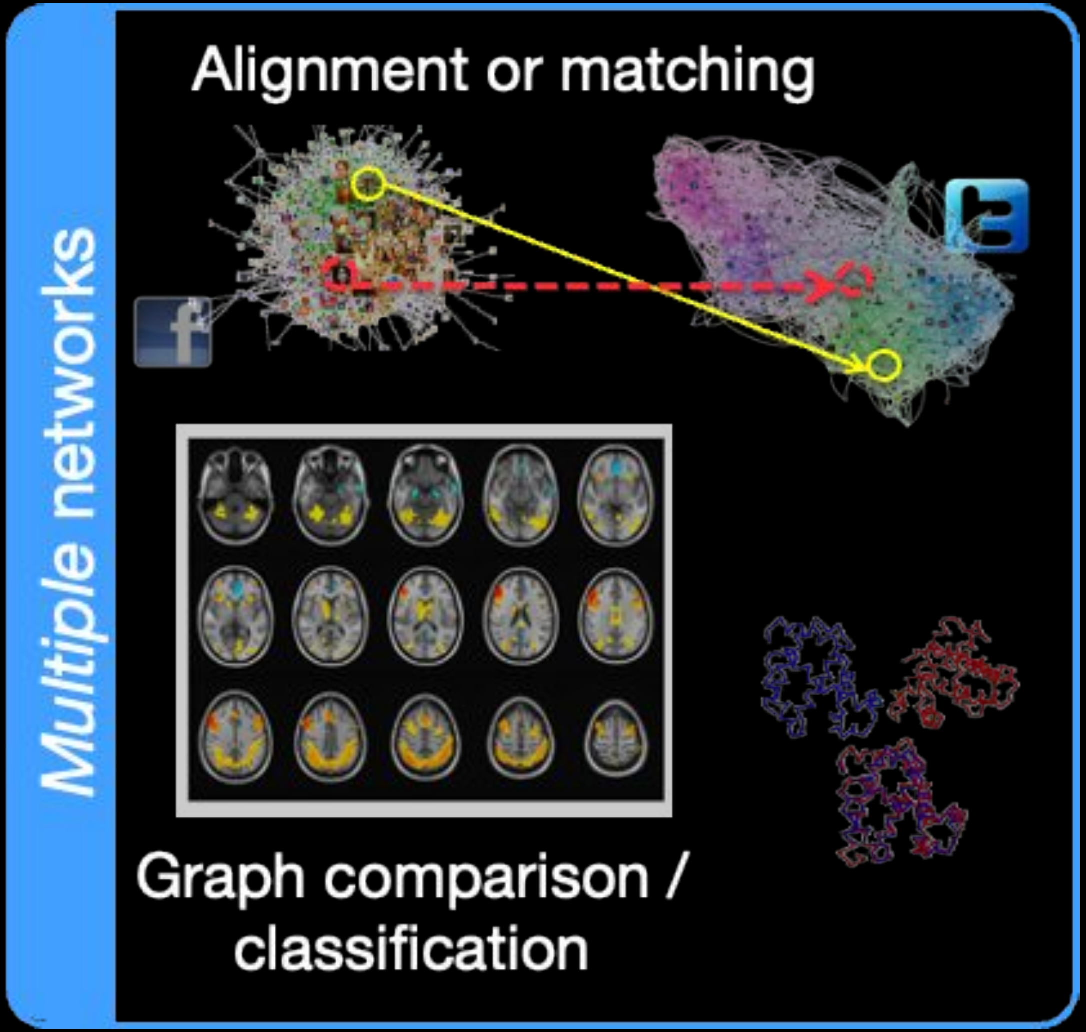


[Henderson et al. KDD'12]

[Lorrain & White '71] [Borgatti & Everett '92] [Wasserman & Faust. '94]

Applications of Structural Role Mining





Tutorial Outline:
Network Embedding for Role Discovery

- Part I: Lecture
  - Introduction
  - **Structural roles in**
    - network science
    - **mathematical sociology**
  - Structural or role-based embedding methods
  - Mining structural roles within a network
  - Mining structural roles across networks
- Part II: Demo
  - Hands-on demo

Let's travel back in time… circa 1971-1976

Journal of Mathematical Sociology
1971, Vol. 1, pp 49–80

STRUCTURAL EQUIVALENCE OF INDIVIDUALS IN SOCIAL NETWORKS†

FRANÇOIS LORRAIN, HARRISON C. WHITE
Harvard University

The aim of this paper is to understand the interrelations among relations within concrete social groups. Social structure is sought, not ideal types, although the latter are relevant to interrelations among relations. From a detailed social network, patterns of global relations can be extracted, within which classes of equivalently positioned individuals are delineated. The global patterns are derived algebraically through a 'functorial' mapping of the original pattern. Such a mapping (essentially a generalized homomorphism) allows systematically for concatenation of effects through the network. The notion of functorial mapping is of central importance in the 'theory of categories,' a branch of modern algebra with numerous applications to algebra, topology, logic. The paper contains analyses of two social networks, exemplifying this approach.

By interrelations among relations is meant the way in which relations among the members of a social system occur in characteristic bundles and how these bundles of relations interlock and determine one another.‡ By understanding is meant distilling simpler patterns at a higher level of abstraction—simpler not only in having fewer constituents but also in exhibiting interrelations which are more regular or transparent.

Practicable ways of carrying out analyses with data have been developed, in the

*Positions in Networks**

RONALD S. BURT, *University of California, Berkeley*

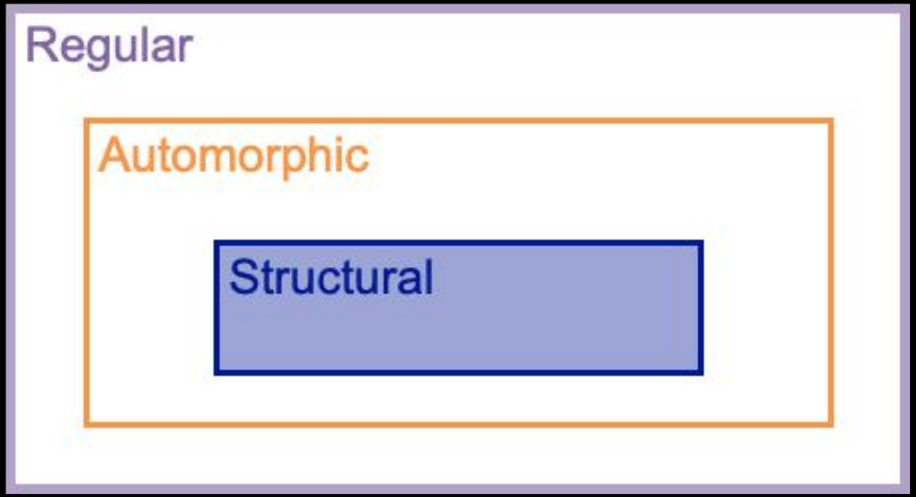
ABSTRACT

The existence of an actor as a set of asymmetric relations to and from every actor in a network of relations is specified as the position of the actor in the network. Conditions of strong versus weak structural equivalence of actor positions in a network are defined. Network structure is characterized in terms of structurally nonequivalent, jointly occupied, network positions located in the observed network. The social distances of actors from network positions are specified as unobserved variables in structural equation models in order to extend the analysis of networks into the etiology and consequences of network structure.

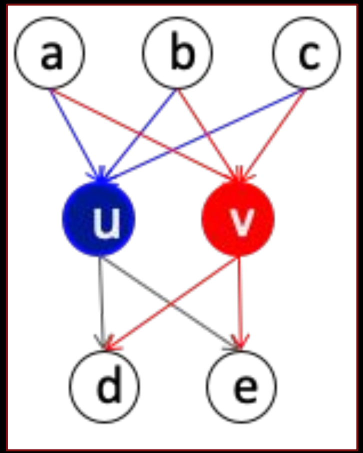
We are each nested in a cacophony of relations with other actors in society. These relations serve to define our existence in society. We are who we are as a function of our relations to and from other actors in society. With the growth of technology and its concomitant division of labor, the determination of actors in society as a function of their relations with other actors is likely to increase rather than decrease. The problem for the social scientist then becomes one of conceptualizing the patterns of relations between an actor and the social system in which he exists in a manner optimally suited to explanation.

Within the total set of all relations which link an actor to other actors in a social system, there are subsets of similar relations. There are economic relations linking the actor to specific other actors. There are relations of friendship, relations of kinship, and relations of status. There are political relations linking the actor to other actors. The list has no end. Each of these types of relations among actors in a social system serves to define a network of relations among the actors. This paper elaborates a conceptualization of networks of relations among actors in a system which simultaneously captures the basic characteristics of the structure in an observed network of relations and easily lends itself to the investigation of the etiology and consequences of that structure through the use of structural equation models.

Deterministic equivalences

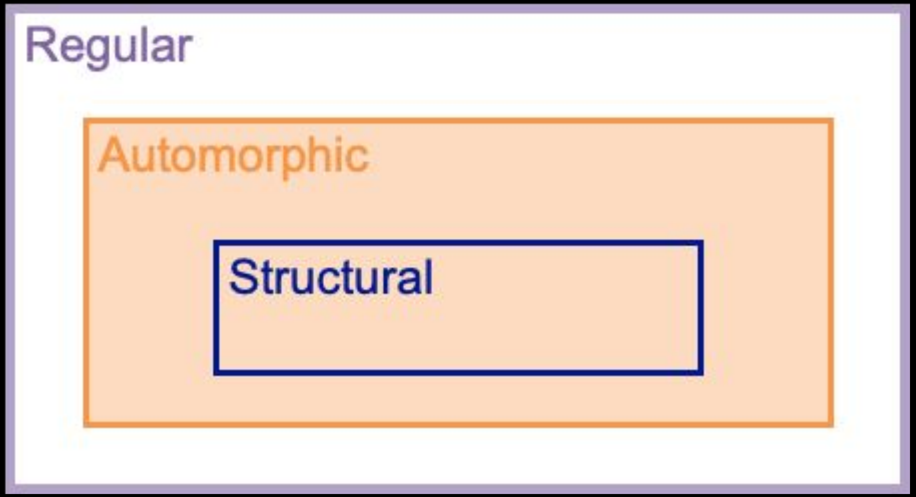


- Nodes u and v are *structurally equivalent* if they have the same relationships to all other nodes
- Rare in real networks

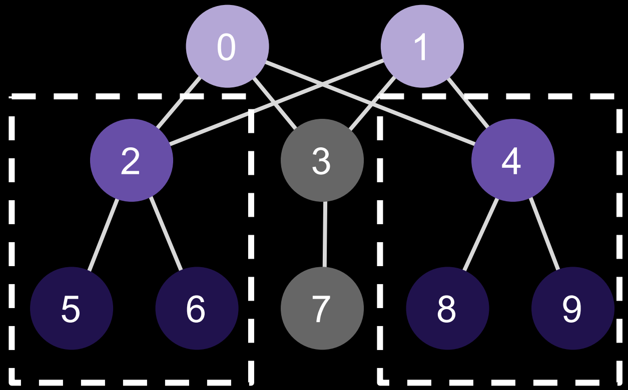


Proximity-based methods tend to capture **structural** equivalence.

Deterministic equivalences

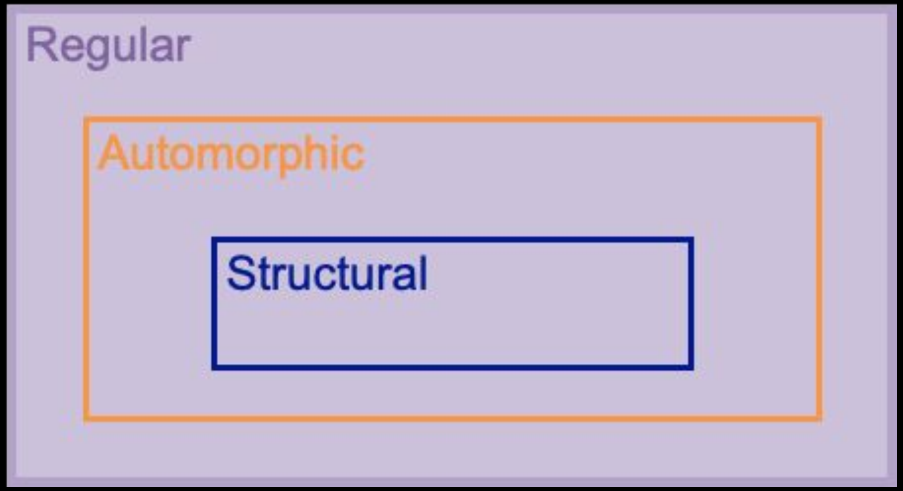


- Nodes u and v are *automorphically equivalent* if all the nodes can be relabeled to form an isomorphic graph with the labels of u and v interchanged
- They share the same label-independent properties

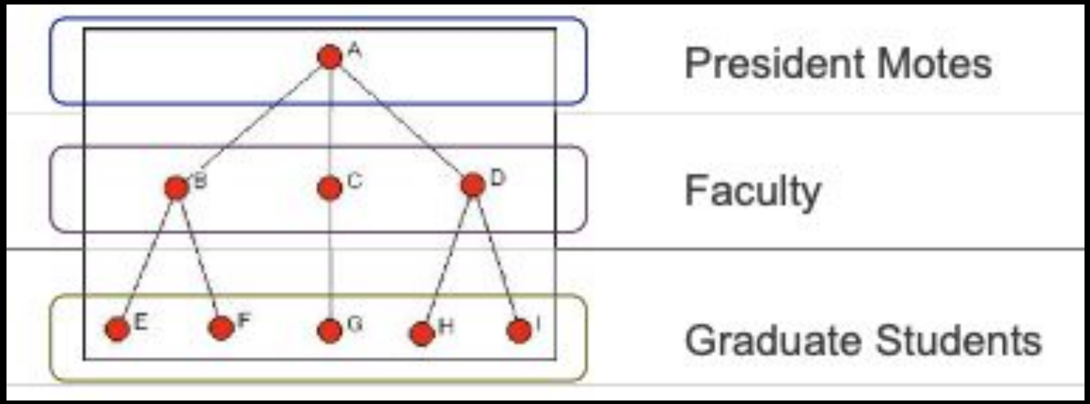


Automorphically Equivalent Groups:
{0, 1} {2, 4} {5, 6, 8, 9}

Deterministic equivalences



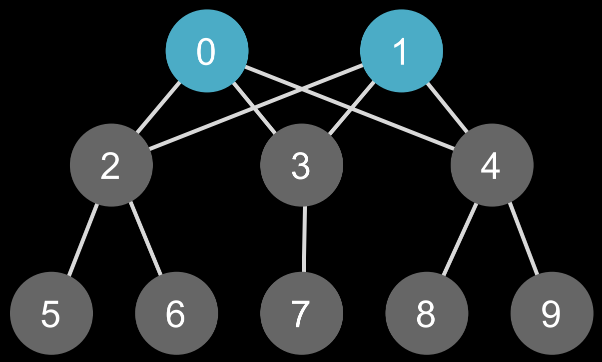
- Nodes *u* and *v* are *regularly equivalent* if they are equally related to equivalent nodes



Sociological Role Equivalence

STRUCTURAL Equivalence

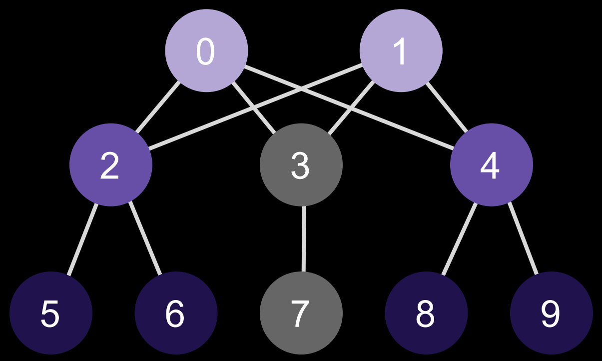
Two nodes are structurally equivalent iff they have identical connections with identical nodes



Structurally Equivalent Group:
{0, 1}

AUTOMORPHIC Equivalence

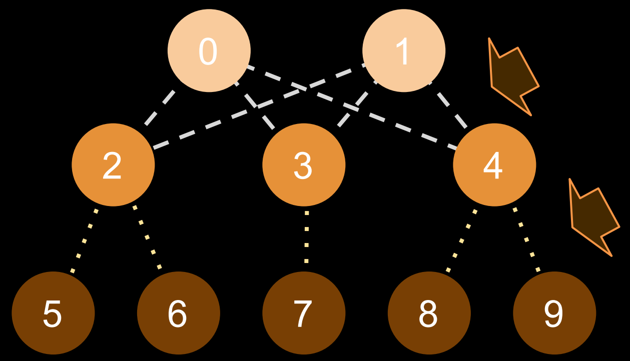
Two nodes are automorphically equivalent iff there is an automorphism that maps one node to the other



Automorphically Equivalent Group:
{0, 1} {2, 4} {5, 6, 8, 9}

REGULAR Equivalence

Two nodes are regularly equivalent if they relate in the same way to equivalent nodes



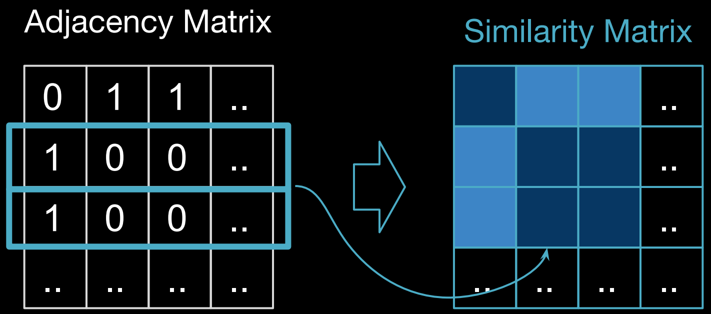
Regularly Equivalent Group:
{0, 1} {2, 3, 4} {5, 6, 7, 8, 9}

UCINET Software for Generating Equivalence

STRUCTURAL Equivalence

Two nodes are structurally equivalent iff they have identical connections with identical nodes

CONCOR [Ronald+ '75]

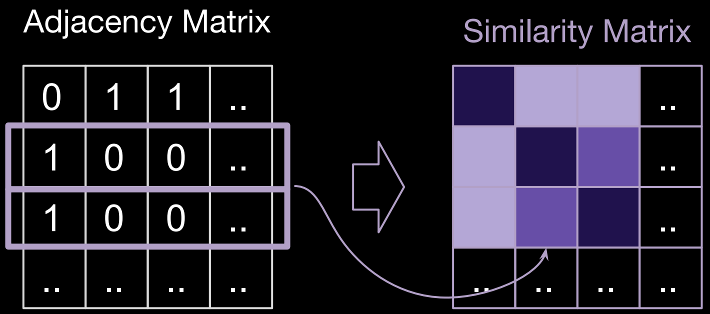


$S_{ij} = S_{ji}$: Pearson correlation between nodes i and j

AUTOMORPHIC Equivalence

Two nodes are automorphically equivalent iff there is an automorphism that maps one node to the other

MAXSIM [Martin+ '88]

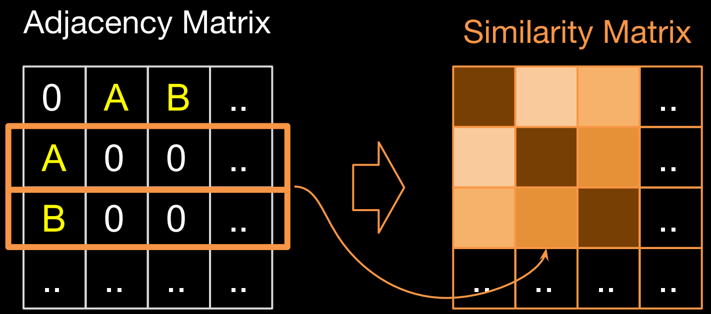


$S_{ij} = S_{ji}$: the similarity of distributions of geodesic distances between nodes i and j to all other nodes

REGULAR Equivalence

Two nodes are regularly equivalent if they relate in the same way to equivalent nodes

CatRege [Stephen+ '92]



$S_{ij} = S_{ji}$: the iteration nodes i and j separated when successively matching node neighborhoods

[Jin, Heimann, Jin, Koutra. 2021. Toward Understanding and Evaluating Structural Node Embeddings. ACM TKDD 2021]

Related Sociology Literature

- S.P. Borgatti and M.G. Everett. 1992. Notions of position in social network analysis. Sociological methodology 22, 1 (1992)
- Stephen P Borgatti, Martin G Everett, and Jeffrey C Johnson. 2018. Analyzing social networks. Sage
- F. Lorrain and H.C. White. 1971. Structural equivalence of individuals in social networks. Journal of Mathematical Sociology
- S. Boorman, H.C. White: Social Structure from Multiple Networks: II. Role Structures. American Journal of Sociology, 81:1384-1446, 1976.
- R.S. Burt: Positions in Networks. Social Forces, 55:93-122, 1976.
- M.G. Everett, S. P. Borgatti: Regular Equivalence: General Theory. Journal of Mathematical Sociology, 19(1):29-52, 1994.
- K. Faust, A.K. Romney: Does Structure Find Structure? A critique of Burt's Use of Distance as a Measure of Structural Equivalence. Social Networks, 7:77-103, 1985.
- K. Faust, S. Wasserman: Blockmodels: Interpretation and Evaluation. Social Networks, 14:5–61. 1992.
- R.A. Hanneman, M. Riddle: Introduction to Social Network Methods. University of California, Riverside, 2005.
- L.D. Sailer: Structural Equivalence: Meaning and Definition, Computation, and Applications. Social Networks, 1:73-90, 1978.
- M.K. Sparrow: A Linear Algorithm for Computing Automorphic Equivalence Classes: The Numerical Signatures Approach. Social Networks, 15:151-170, 1993.
- S. Wasserman, K. Faust: Social Network Analysis: Methods and Applications. Cambridge University Press, 1994.
- H.C. White, S. A. Boorman, R. L. Breiger: Social Structure from Multiple Networks I. Blockmodels of Roles and Positions. American Journal of Sociology, 81:730-780, 1976.
- D.R. White, K. Reitz: Graph and Semi-Group Homomorphism on Networks and Relations. Social Networks, 5:143-234, 1983.
- …

Tutorial Outline:

Network Embedding for Role Discovery

- Part I: Lecture
  - Introduction
  - Structural roles in
    - network science
    - mathematical sociology
  - **Structural or role-based embedding methods**
  - Mining structural roles within a network
  - Mining structural roles across networks
- Part II: Demo
  - Hands-on demo

A lot of work on network representation learning

Must-read papers on NRL/NE.

NRL: network representation learning. NE: network embedding.

Contributed by Cunchao Tu, Yuan Yao and Zhengyan Zhang.

We release OpenNE, an open source toolkit for NE/NRL. This repository p
Representation Learning) training and testing framework. Currently, the
DeepWalk, LINE, node2vec, GraRep, TADW and GCN.

Survey papers:

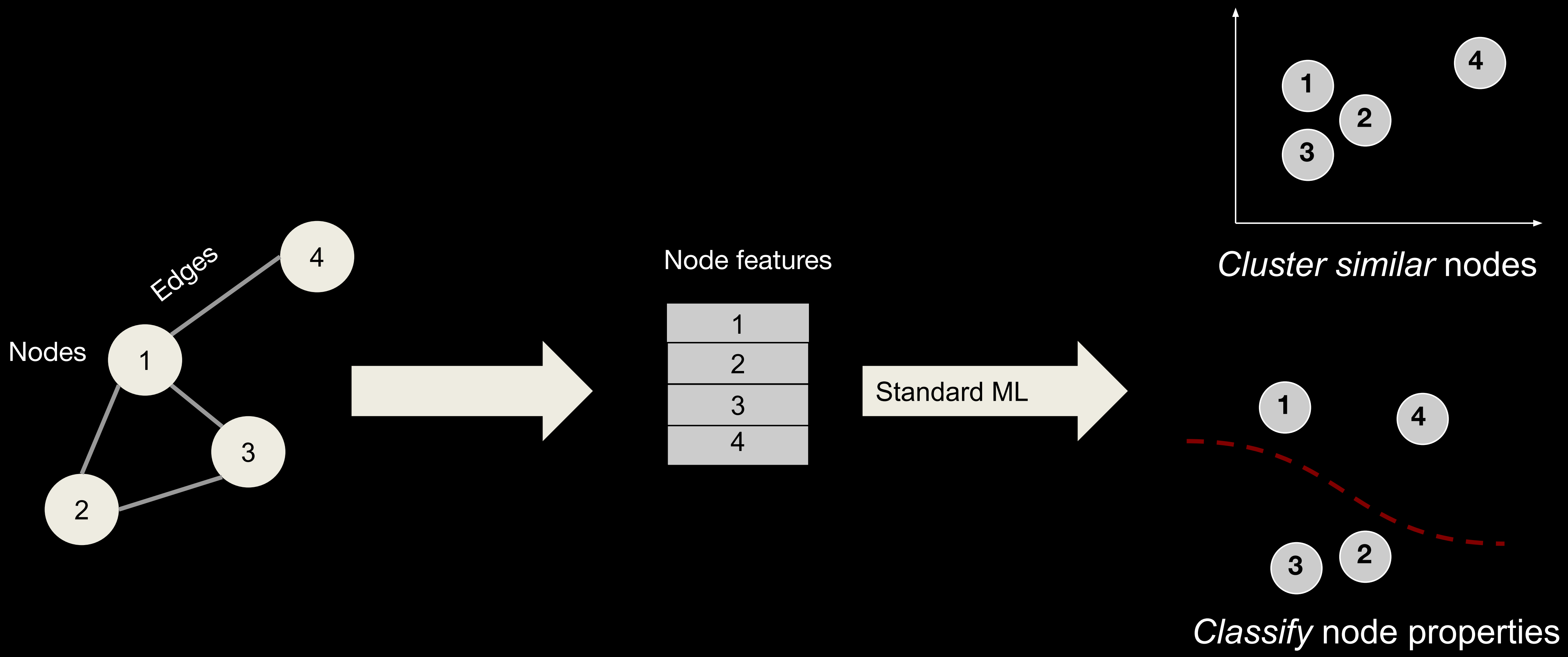
1. **Representation Learning on Graphs: Methods and Applications.** W
2017. paper
2. **Graph Embedding Techniques, Applications, and Performance:** A
3. **A Comprehensive Survey of Graph Embedding: Problems, Techni**
Zheng, Kevin Chen-Chuan Chang. 2017. paper
4. **Network Representation Learning: A Survey.** *Daokun Zhang, Jie Yi*
5. **A Tutorial on Network Embeddings.** *Haochen Chen, Bryan Perozzi,*
6. **Network Representation Learning: An Overview.(In Chinese)** *Cunc*
2017. paper
7. ***Relational inductive biases, deep learning, and graph networks.*** *P*
Bapst, Alvaro Sanchez-Gonzalez, Vinicius Zambaldi, Mateusz Malino
Santoro, Ryan Faulkner, Caglar Gulcehre, Francis Song, Andrew Balla
Kelsey Allen, Charles Nash, Victoria Langston, Chris Dyer, Nicolas He

54. **Link Prediction via Subgraph Embedding-Based Convex Matrix Completion.** *Zhu Cao, Linlin Wang, Gerard De melo.* AAAI 2018.
55. **Generative Adversarial Network based Heterogeneous Bibliographic Network Representation for Personalized Citation Recommendation.** *J. Han, Xiaoyan Cai, Libin Yang.* AAAI 2018.
56. **DepthLGP: Learning Embeddings of Out-of-Sample**
Zhu. AAAI 2018. paper
57. **Structural Deep Embedding for Hyper-Networks.** *Ke*
paper
58. **TIMERS: Error-Bounded SVD Restart on Dynamic N**
Zhu. AAAI 2018. paper
59. **Community Detection in Attributed Graphs: An Emb**
Zhang. AAAI 2018.
60. **Bernoulli Embeddings for Graphs.** *Vinith Misra, Sumi*
61. **Distance-aware DAG Embedding for Proximity Sear**
Zhou Zhao, Fanwei Zhu, Kevin Chen-Chuan Chang, M
62. **GraphGAN: Graph Representation Learning with Ge**
Wang, MIAO ZHAO, Weinan Zhang, Fuzheng Zhang, X
63. **HARP: Hierarchical Representation Learning for Net**
AAAI 2018. paper code
64. **Representation Learning for Scale-free Networks.** R
2018. paper
65. **Social Rank Regulated Large-scale Network Embed**
2018. paper

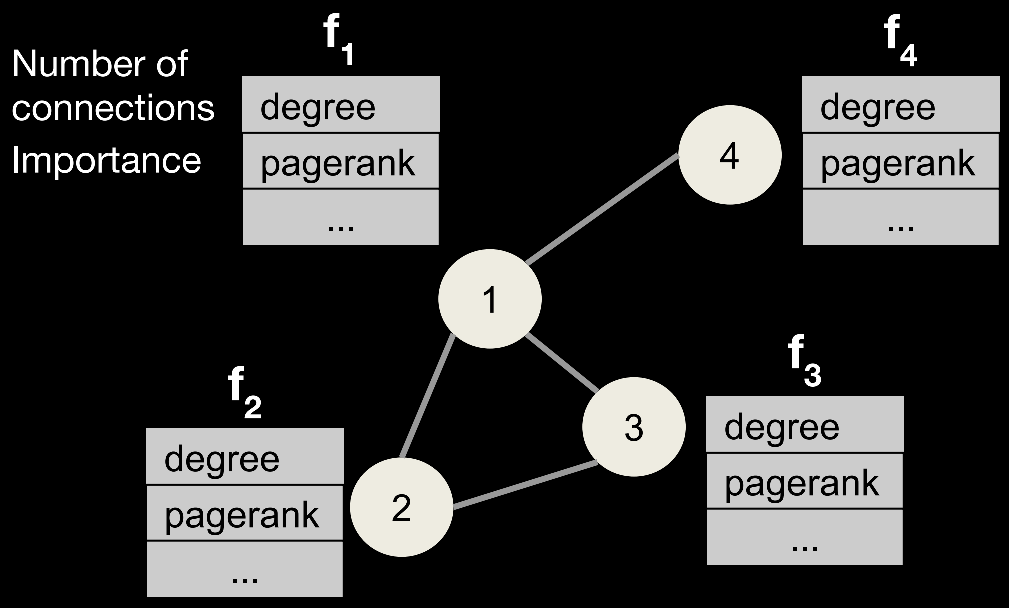
101. **Integrative Network Embedding via Deep Joint Reconstruction.** *Di Jin, Meng Ge, Liang Yang, Dongxiao He, Longbiao Wang, Weixiong Zhang.* IJCAI 2018.
102. **Scalable Multiplex Network Embedding.** *Hongming Zhang, Liwei Qiu, Lingling Yi, Yangqiu Song.* IJCAI 2018. paper
103. **Adversarially Regularized Graph Autoencoder for Graph Embedding.** *Shirui Pan, Ruiqi Hu, Guodong Long, Jing Jiang, Lina Yao, Chengqi Zhang.* IJCAI 2018.
104. **Dynamic Network Embedding : An Extended Approach for Skip-gram based Network Embedding.** *Lun Du, Yun Wang, Guojie Song, Zhicong Lu, Junshan Wang.* IJCAI 2018.
105. **Discrete Network Embedding.** *Xiaobo Shen, Shirui Pan, Weiwei Liu, Yew-Soon Ong, Quan-Sen Sun.* IJCAI 2018.
106. **Deep Attributed Network Embedding.** *Hongchang Gao, Heng Huang.* IJCAI 2018.
107. **Active Discriminative Network Representation Learning.** *Li Gao, Hong Yang, Chuan Zhou, Jia Wu, Shirui Pan, Yue Hu.* IJCAI 2018.
108. **ANRL: Attributed Network Representation Learning via Deep Neural Networks.** *Zhen Zhang, Hongxia Yang, Jiajun Bu, Sheng Zhou, Pinggang Yu, Jianwei Zhang, Martin Ester, Can Wang.* IJCAI 2018.
109. **Feature Hashing for Network Representation Learning.** *Qixiang Wang, Shanfeng Wang, Maoguo Gong, Yue Wu.* IJCAI 2018.
110. **Constructing Narrative Event Evolutionary Graph for Script Event Prediction.** *Zhongyang Li, Xiao Ding, Ting Liu.* IJCAI 2018. paper code
111. **Deep Inductive Network Representation Learning.** *Ryan A. Rossi, Rong Zhou, Nesreen K. Ahmed.* WWW 2018. paper
112. **A Unified Framework for Community Detection and Network Representation Learning.** *Cunchao Tu, Xiangkai Zeng, Hao Wang, Zhengyan Zhang, Zhiyuan Liu, Maosong Sun, Bo Zhang, Leyu Lin.* TKDE 2018. paper

...

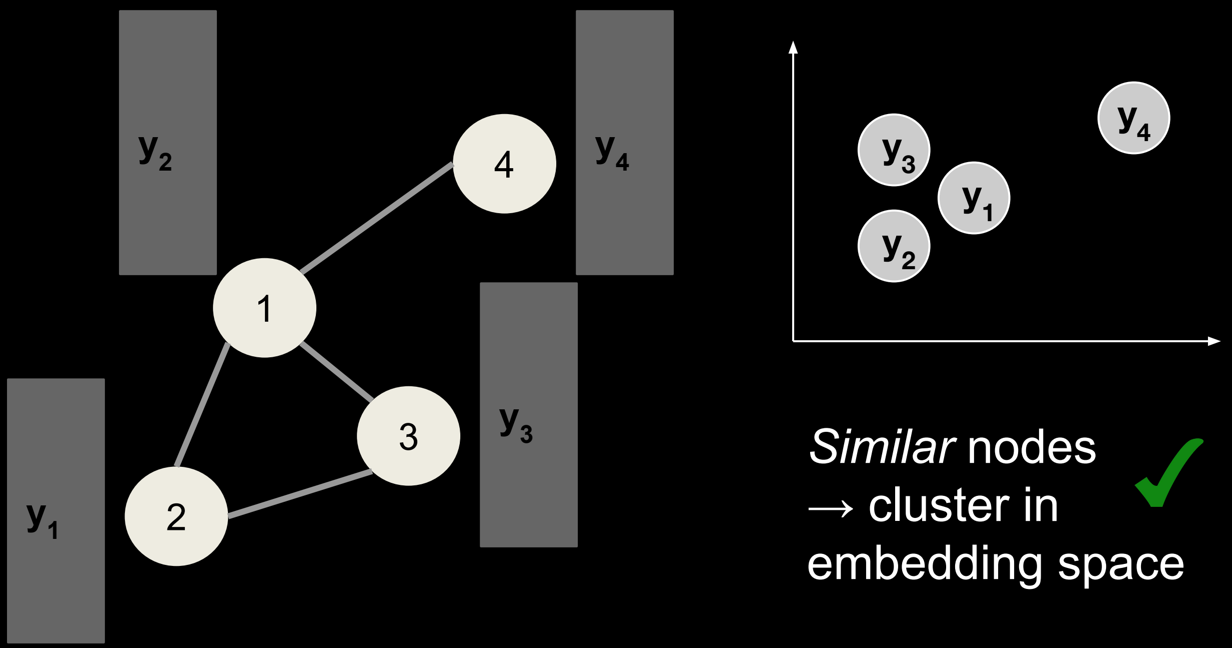
Learning with Graphs



How to Get Node Features?



Traditional Approaches:
Hand-Engineered Features
- Interpretable ✓
- Simplistic, hard to select ✗

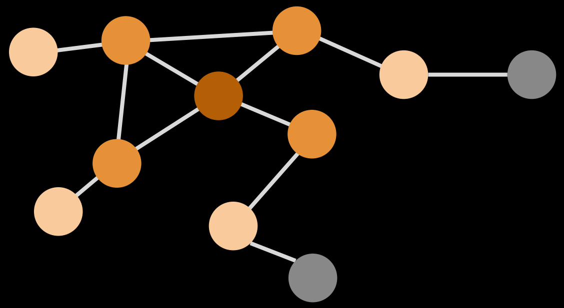


Similar nodes → cluster in embedding space ✓

Node Embeddings [Perozzi+ '14], [Tang+ '15], [Grover+ '16], [Ribeiro+ '17], ...
- Latent features
- Preserve complex **similarity**

LINE

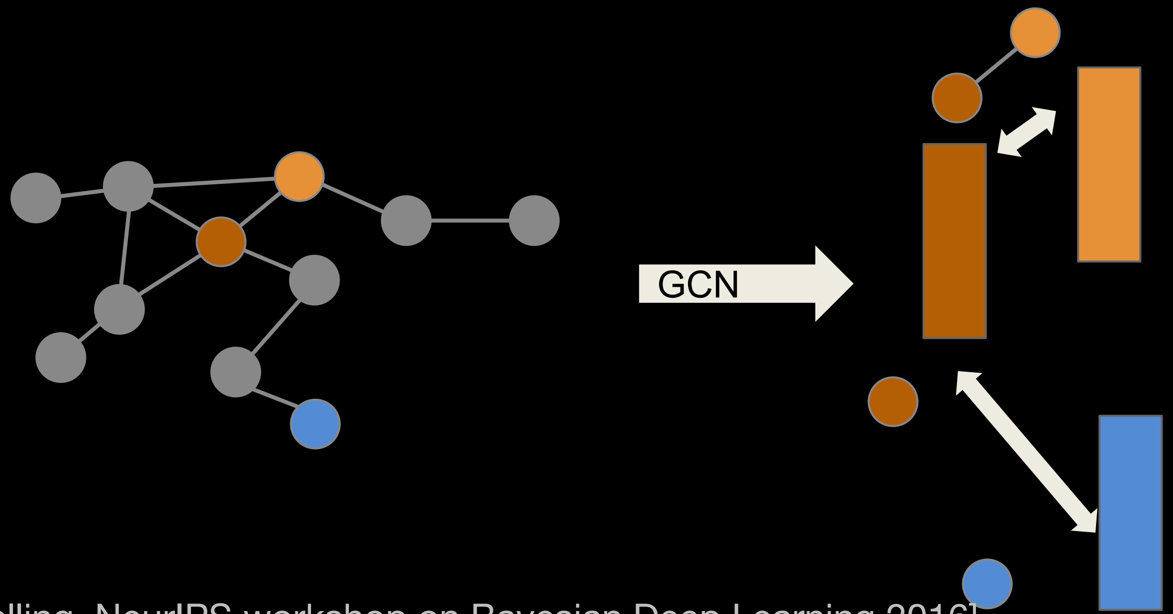
- Primarily model node proximity rather than structural roles
- Embedding objective: learn similar representations for first and second order neighbors



[Tang, Qu, et al. WWW, 2015]

GCN-VAE

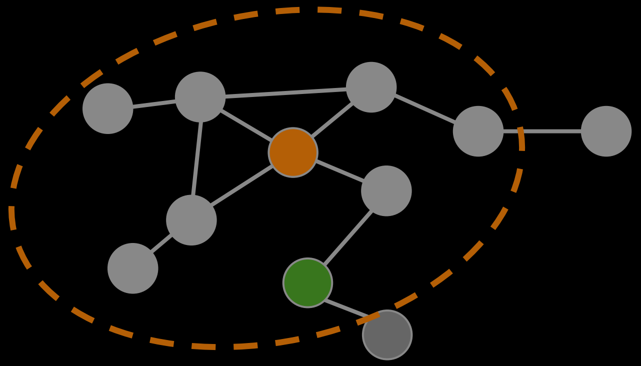
- Use graph convolutional network to learn node feature vectors
- *Autoencoder* paradigm: training objective is for features to reconstruct graph structure (similar features = nodes share an edge)



[Kipf & Welling. NeurIPS workshop on Bayesian Deep Learning 2016]

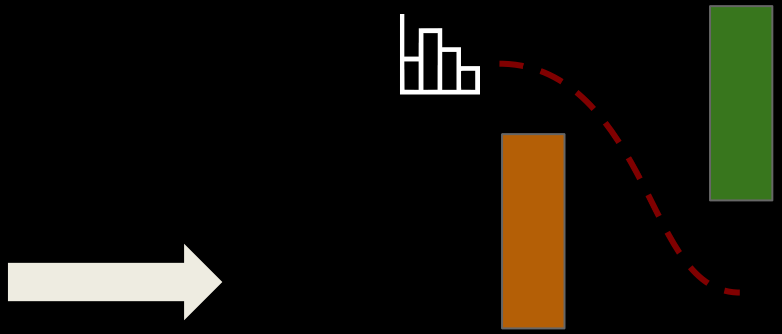
xNetMF

- Characterize connectivity statistics of local neighborhood
- Embedding objective: similar embeddings for similar neighborhoods



Degree histogram of
k-hop neighbors

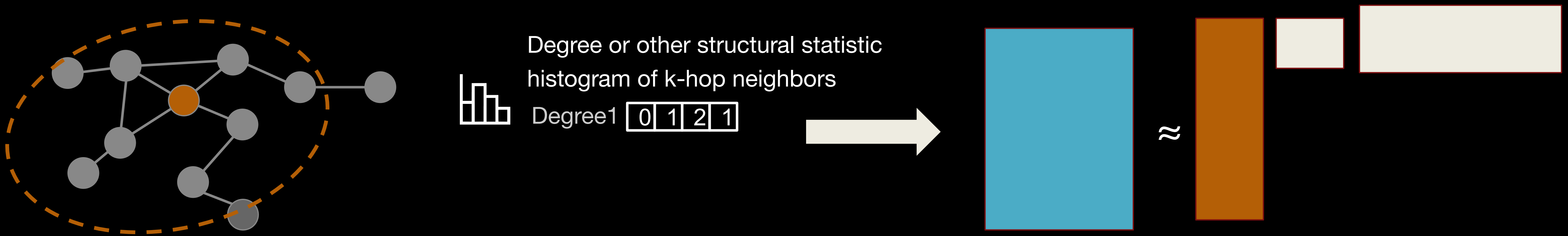
Degree1	0	1	2	1



[Heimann, Shen, et al. CIKM 2018]

MultiLENS

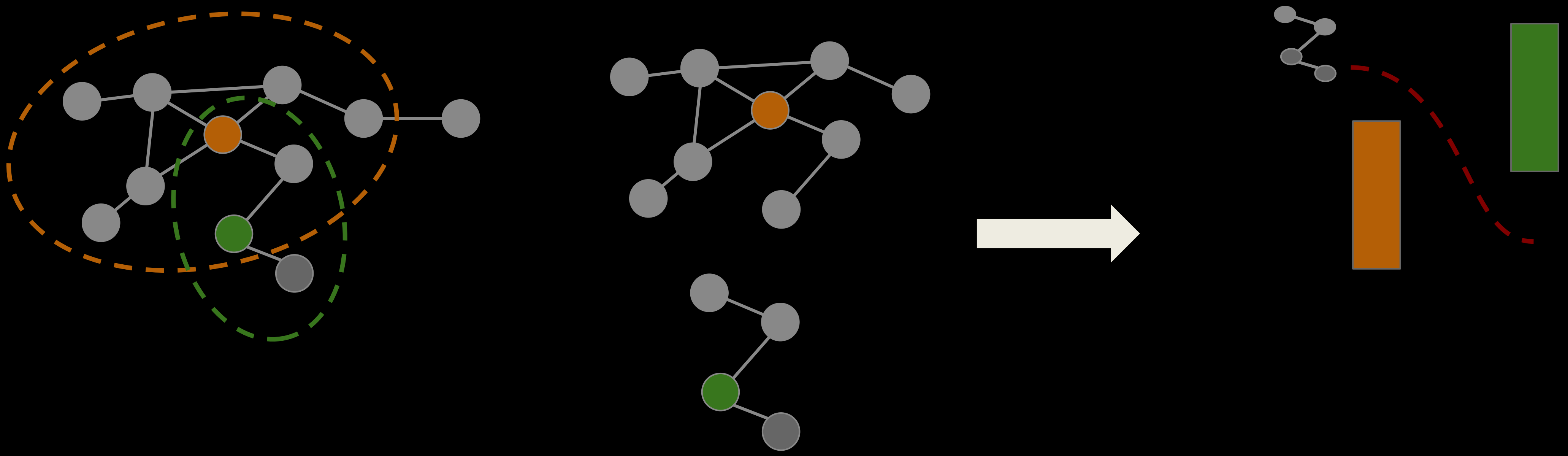
- Characterize distribution of structural statistics of local neighborhood
- Embedding: Low rank decomposition of feature matrix



[Jin, Rossi, et al. KDD 2019]

SEGK

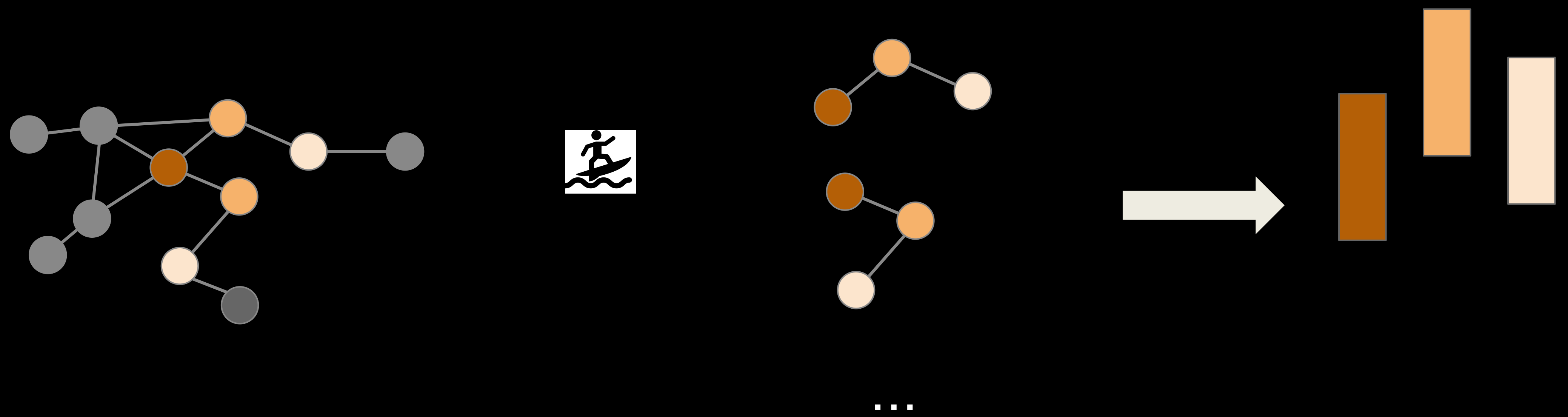
- Extract local neighborhood around each node
- Characterize local neighborhood using graph kernels
- Embedding objective: similar embeddings for similar neighborhoods



[Nikolentzos and Vazirgiannis. TKDE 2019]

node2vec

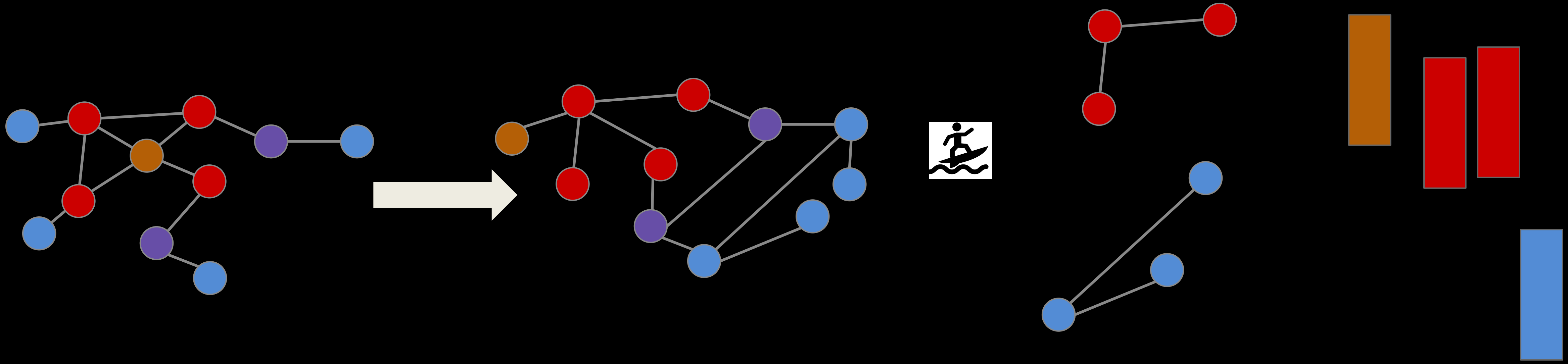
- Perform random walks on graph
- Embedding objective: similar embeddings for nodes that co-occur in random walks



[Grover and Leskovec. KDD 2016]

struc2vec

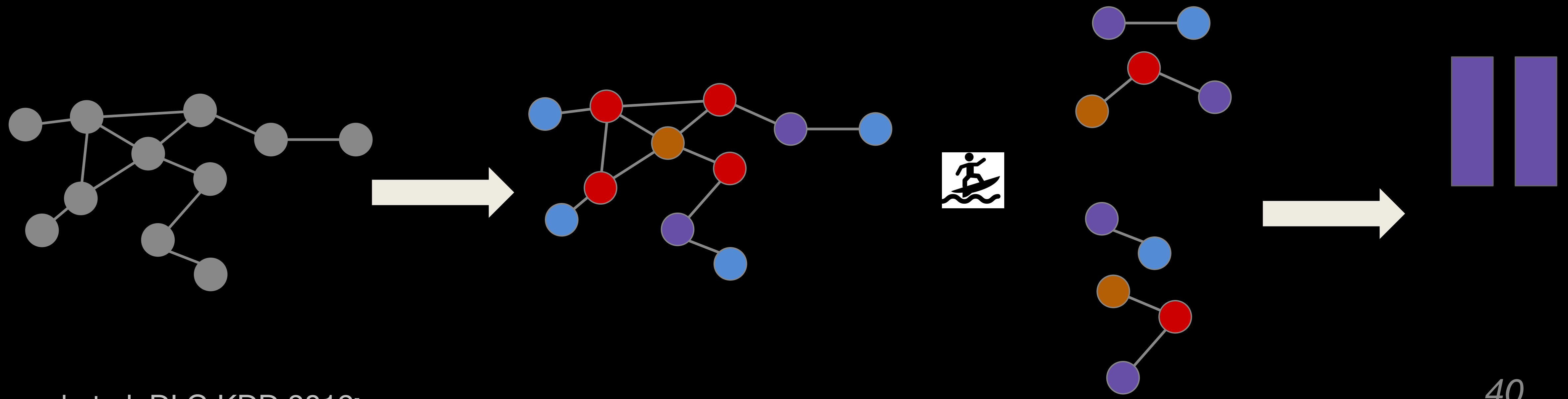
- Perform random walks on structural similarity graph
- Structural similarity determined by comparing neighborhood connectivity statistics at multiple levels
- Same embedding objective: similar embeddings for nodes that co-occur in random walks



[Ribeiro et al. KDD 2017]

role2vec

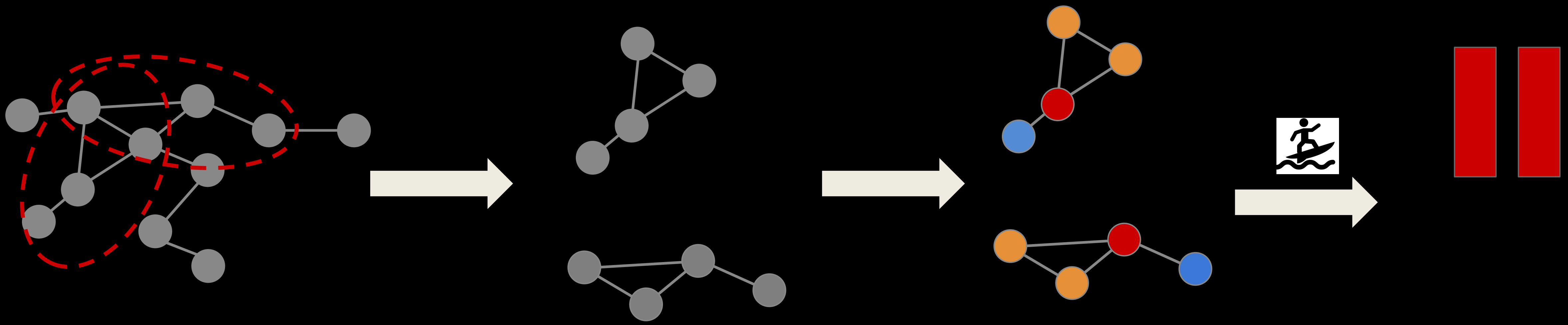
- Relabel nodes by structural role
- Perform random walks on original graph
- Embedding objective: embed nodes similarly that co-occur with similar types



[Ahmed et al. DLG KDD 2019]

RiWalk

- Extract subgraph around each node
- Relabel structural positions of nodes in each subgraph
- Perform random walks on subgraph, same embedding objective



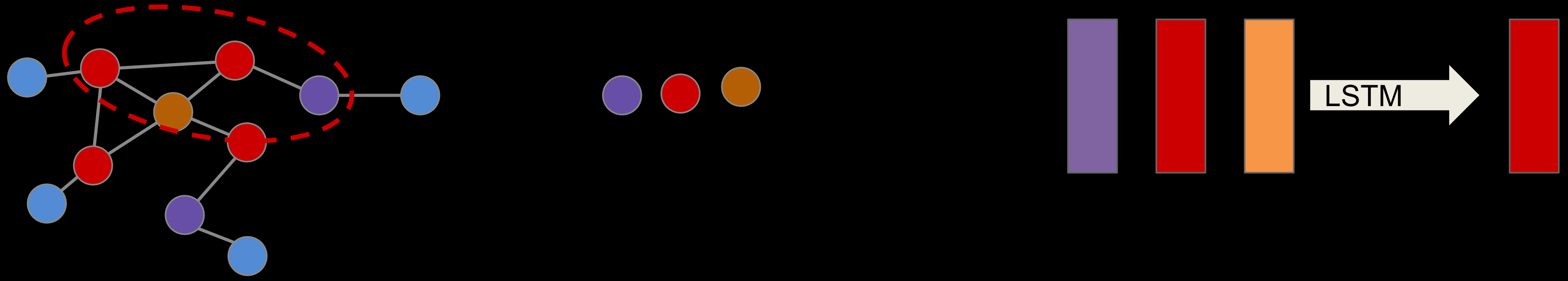
[Xuewei et al.. ICDM 2019]

DRNE

Sort neighborhoods by degree

Aggregate neighbors' embeddings using LSTM

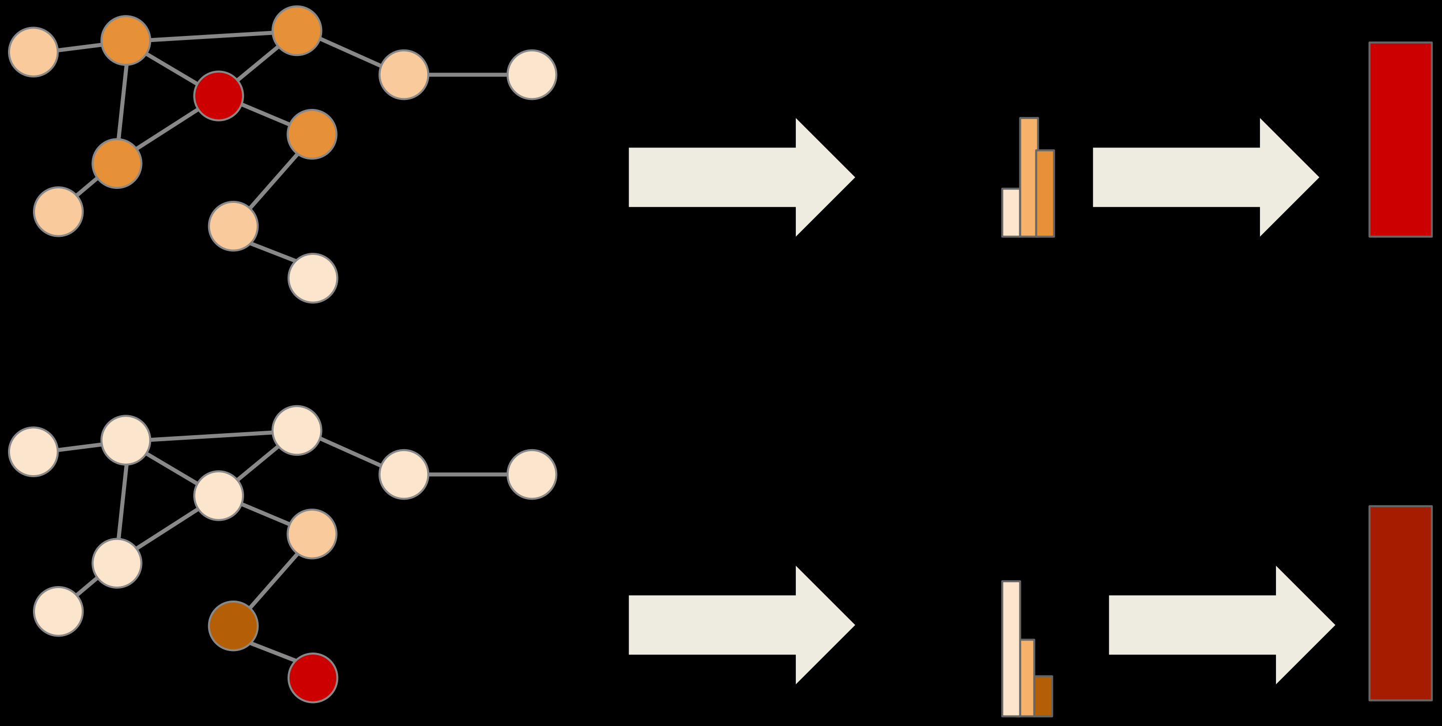
- Additional regularization so that embedding approximates node degree
- Claims to have some power to model regular equivalence



[Tu et al.. KDD 2018]

GraphWave

- Perform heat diffusion on graph
- Node features = shape of heat distribution sent to other nodes



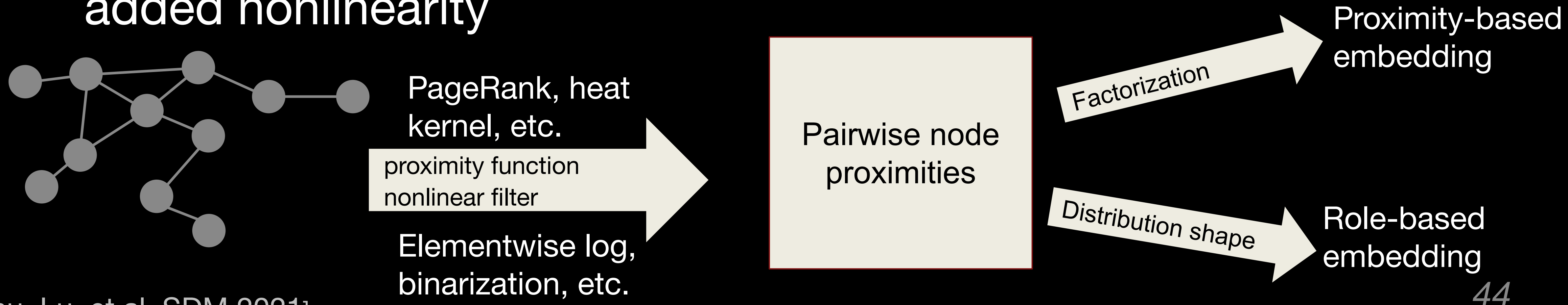
[Donnat et al. KDD 2018]

Phusion: Unifying Role and Proximity-based Embeddings

Role-based and proximity-based methodologies different? [Rossi et. al TKDD 2021] Or similar? [ICLR 2020]

PhUSION: construct unified framework for proximity-preserving and role-based embedding

- Allows for sharing of design choices like proximity function, added nonlinearity



[Zhu, Lu, et al. SDM 2021]

(maybe add references to more structural embedding methods)

Tutorial Outline:
Network Embedding for Role Discovery

- Part I: Lecture
  - Introduction
  - Structural roles in
    - network science
    - mathematical sociology
  - Structural or role-based embedding methods

  - Mining structural roles within a network
  - Mining structural roles across networks
- Part II: Demo
  - Hands-on demo

Questions so far??

Tutorial Outline:

Network Embedding for Role Discovery

- Part I: Lecture
  - Introduction
  - Structural roles in
    - network science
    - mathematical sociology
  - Structural or role-based embedding methods
  - **Mining structural roles within a network**
  - Mining structural roles across networks
- Part II: Demo
  - Hands-on demo

Connecting Network Embedding & Sociology

STRUCTURAL Equivalence

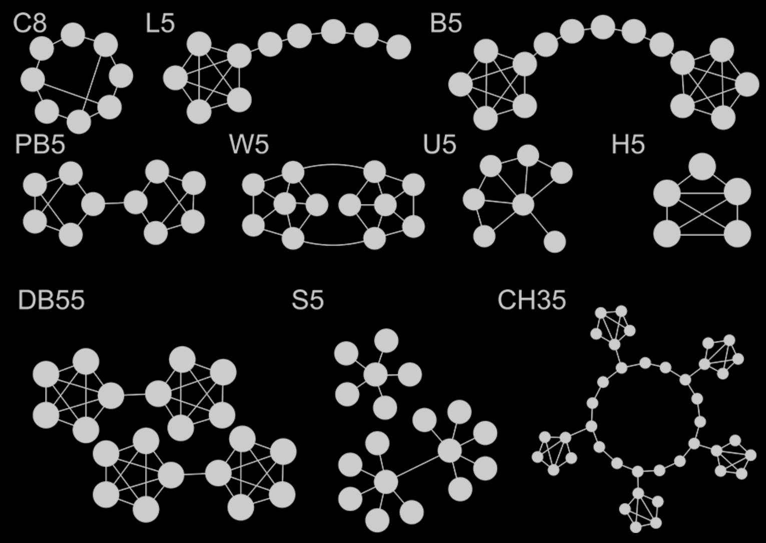
Identical relationships to all other nodes

AUTOMORPHIC Equivalence

Structure-preserving mapping between nodes

REGULAR Equivalence

Equivalent relationships to equivalent other nodes



Synthetic Datasets

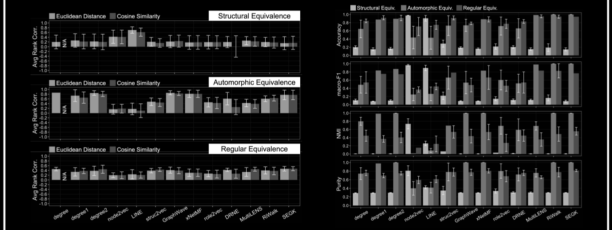
- Air Traffic
- Protein
- Blog
- Facebook
- Email
- ...

Real Datasets

- node2vec/LINE
- struc2vec
- xNetMF
- DRNE
- RiWalk
- GCN-VAE
- GraphWave
- role2vec
- MultiLENS
- SEGK

Structural Embedding Methods

INTRINSIC | EXTRINSIC



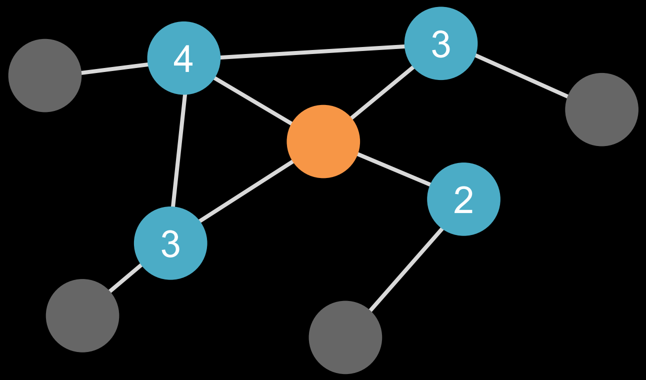
Evaluation

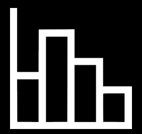
[Jin, Heimann, Jin, Koutra. 2021. Toward Understanding and Evaluating Structural Node Embeddings. ACM TKDD 2021]

Additional Structural “Embedding” Method: Degree Histograms

Degree-*k*: degree histogram of k-hop neighbors

- Degree, Degree1, Degree2 variants





Degree1

0	1	2	1

0 neighbors of degree 1
1 neighbor of degree 2
2 neighbors of degree 3
1 neighbor of degree 4

[Jin, Heimann, Jin, Koutra. 2021. Toward Understanding and Evaluating Structural Node Embeddings. ACM TKDD 2021]

Structural Embedding Graph Library

https://github.com/GemsLab/StrucEmbedding-GraphLibrary

We'll use this graph library during the hands-on part of this tutorial!

README.md

The Structural EMBedding graph library (SEMB)

Authors: GEMS Lab Team @ University of Michigan (Mark Jin, Ruowang Zhang, Mark Heimann)

This SEMB library allows fast onboarding to get and evaluate structural node embeddings. With the unified API interface and the modular codebase, SEMB library enables easy intergration of 3rd-party methods and datasets.

The library itself has already included a set of popular methods and datasets ready for immediate use.

- Built-in methods: node2vec, struc2vec, GraphWave, xNetMF, role2vec, DRNE, MultiLENS, RiWalk, SEGK, (more methods to add in the near future)
- Built-in datasets:

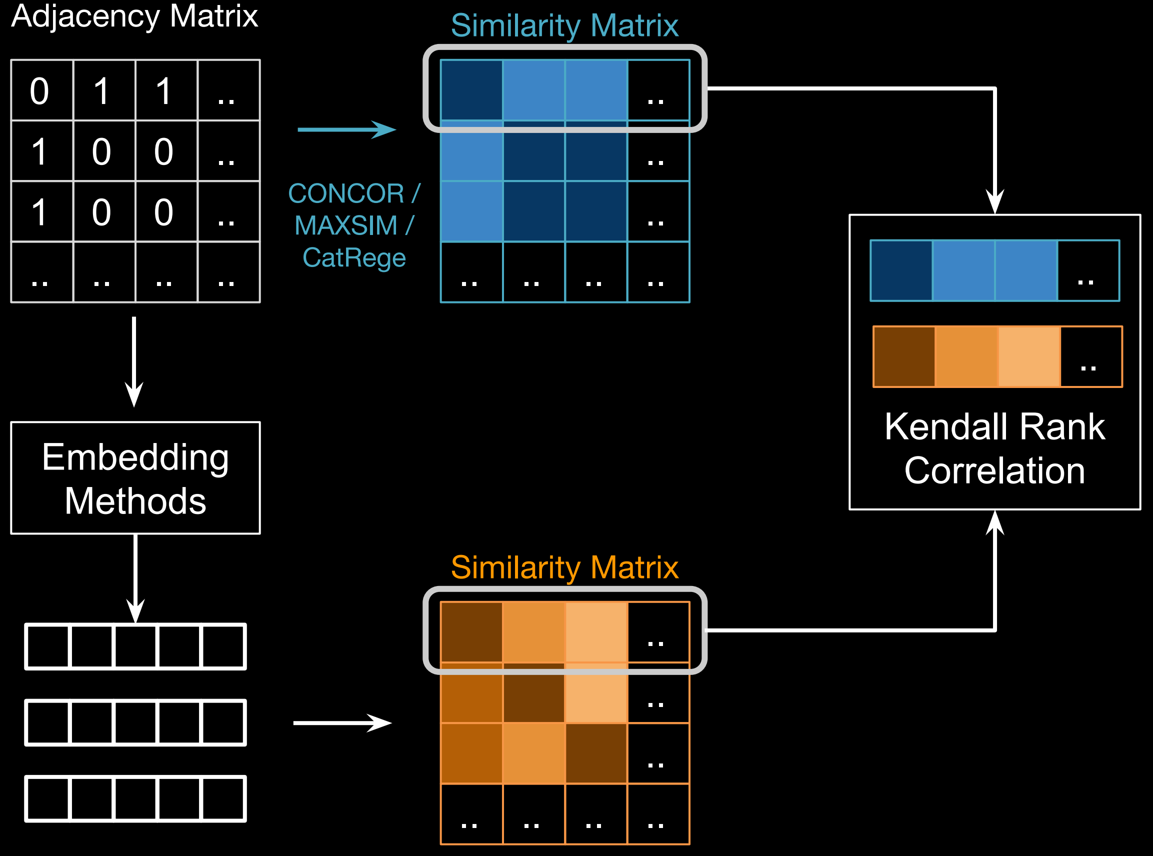
Dataset	# Nodes	# Edges
BlogCatalog	10,312	333,983
Facebook	4,039	88,234
ICEWS	1,255	1,414
PPI	56,944	818,786
BR air-traffic	131	1,038
EU air-traffic	399	5,995
US air-traffic	1,190	13,599
DD6	4,152	20,640
Synthetic Datasets		

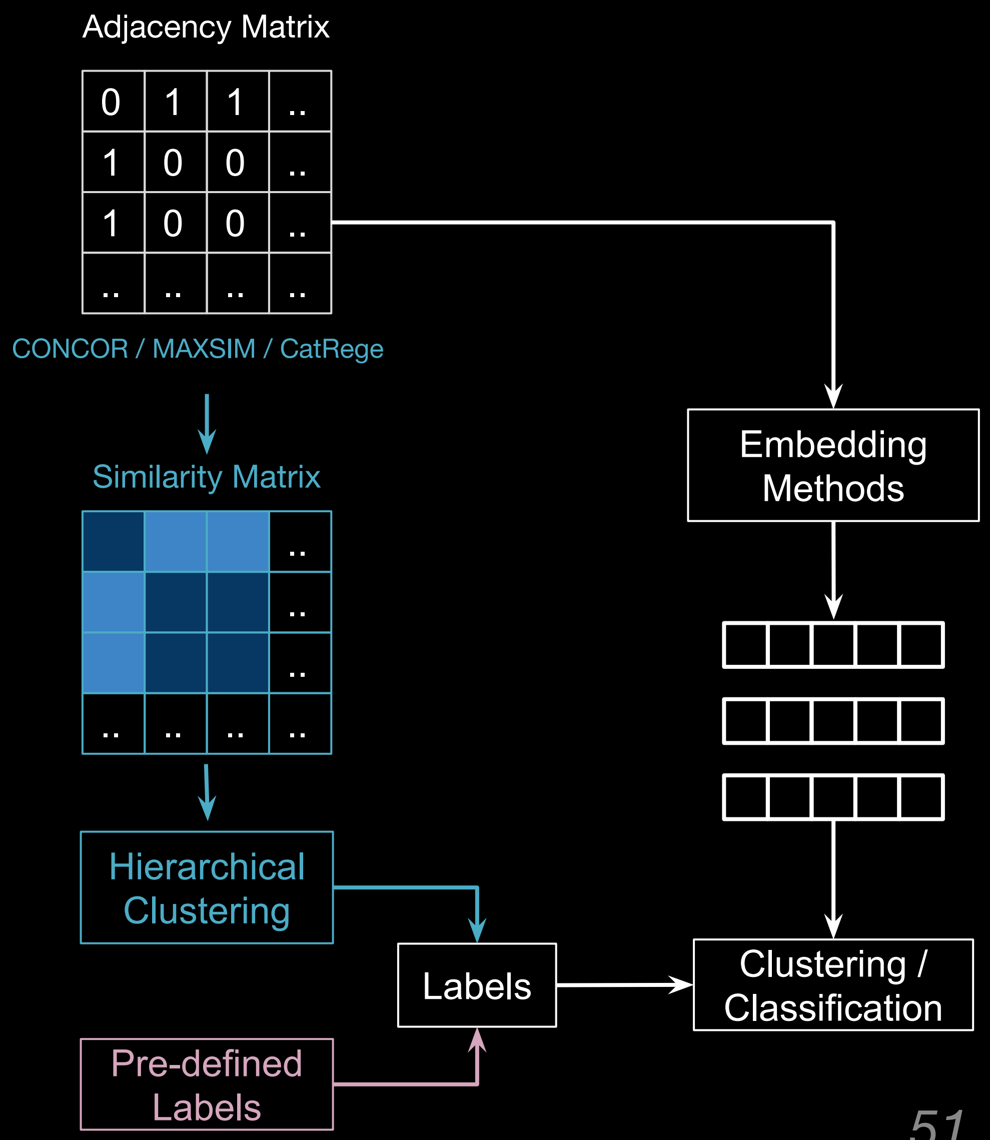
The library requires *Python 3.6.2 for best usage. In *Python 3.8*, the Tensorflow 1.14.0 used in DRNE might not be successfully installed.

Installation and Usage

Make sure you are using *Python 3.6+* for all below!

Intrinsic and Extrinsic Evaluation

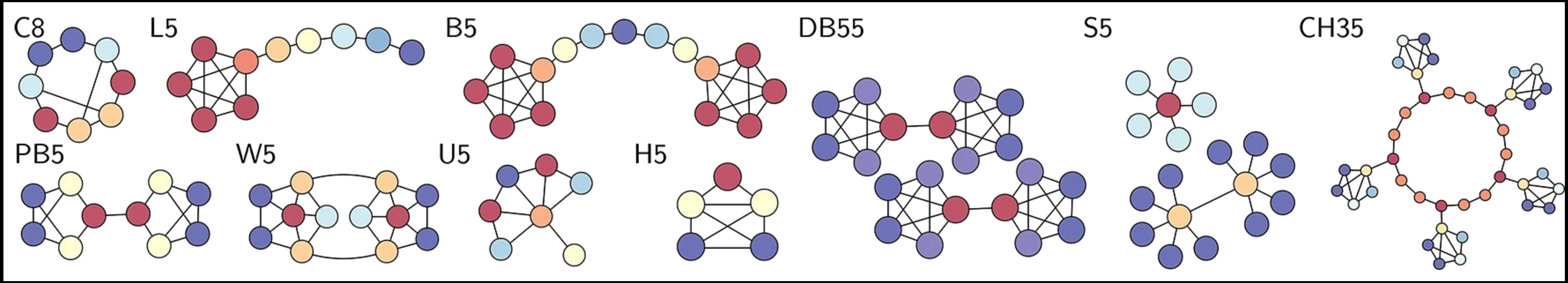




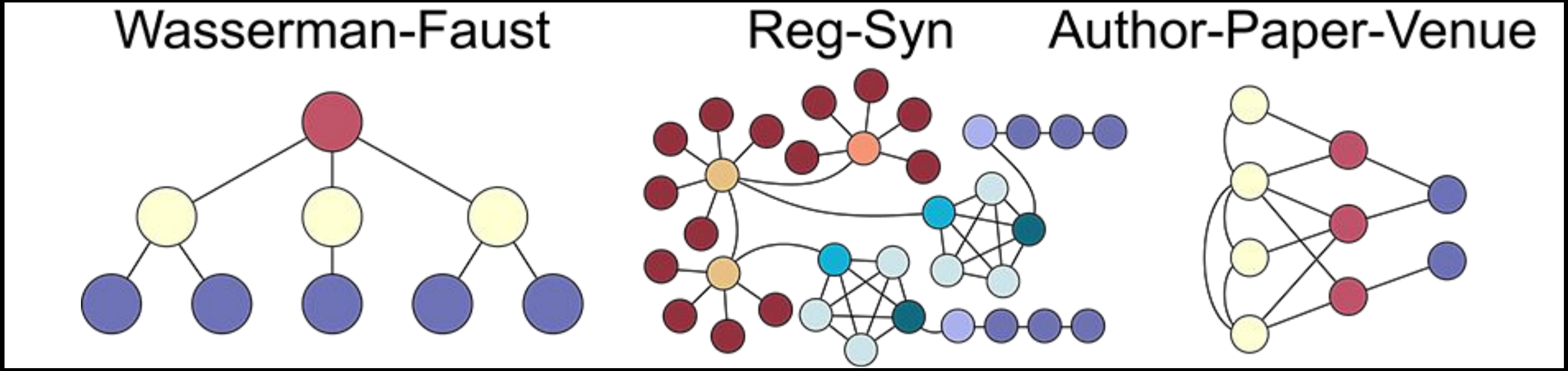
Intrinsic Evaluation doesn’t involve any downstream machine learning model

[Jin, Heimann, Jin, Koutra. 2021. Toward Understanding and Evaluating Structural Node Embeddings. ACM TKDD 2021]

Synthetic Datasets: Base



Identically colored nodes are *automorphically* equivalent



Identically colored nodes are *regularly* equivalent

[Jin, Heimann, Jin, Koutra. 2021. Toward Understanding and Evaluating Structural Node Embeddings. ACM TKDD 2021]

Building Complex Synthetic Benchmarks

Large Graph	Base	Generation
H10_S_L	**H5**	10 H5 on a circle with 2 circular nodes between each connecting circular node with house's side.
H10_T_L	**H5**	10 H5 on a circle with 2 circular nodes between each connecting circular node with house's roof.
Barbell L-A	**B5**	Connecting the out-most nodes on the chain of B5 into a circle.
Barbell L-B	**B5**	Connecting the out-most nodes on the chain of B5 into a circle. Additional 5-clique at each connector.
Ferris Wheel	**C8**	Enlarged version of C8 with similar perturbation.
City of Stars	**S5**	10 normal stars and 5 binary stars as in S5
PB-L	**PB5**	10 half-sided PB5 connected to each node of a 10-node circular graph. All the node degrees are 3.

[Jin, Heimann, Jin, Koutra. 2021. Toward Understanding and Evaluating Structural Node Embeddings. ACM TKDD 2021]

Real Datasets: Single Network Mining

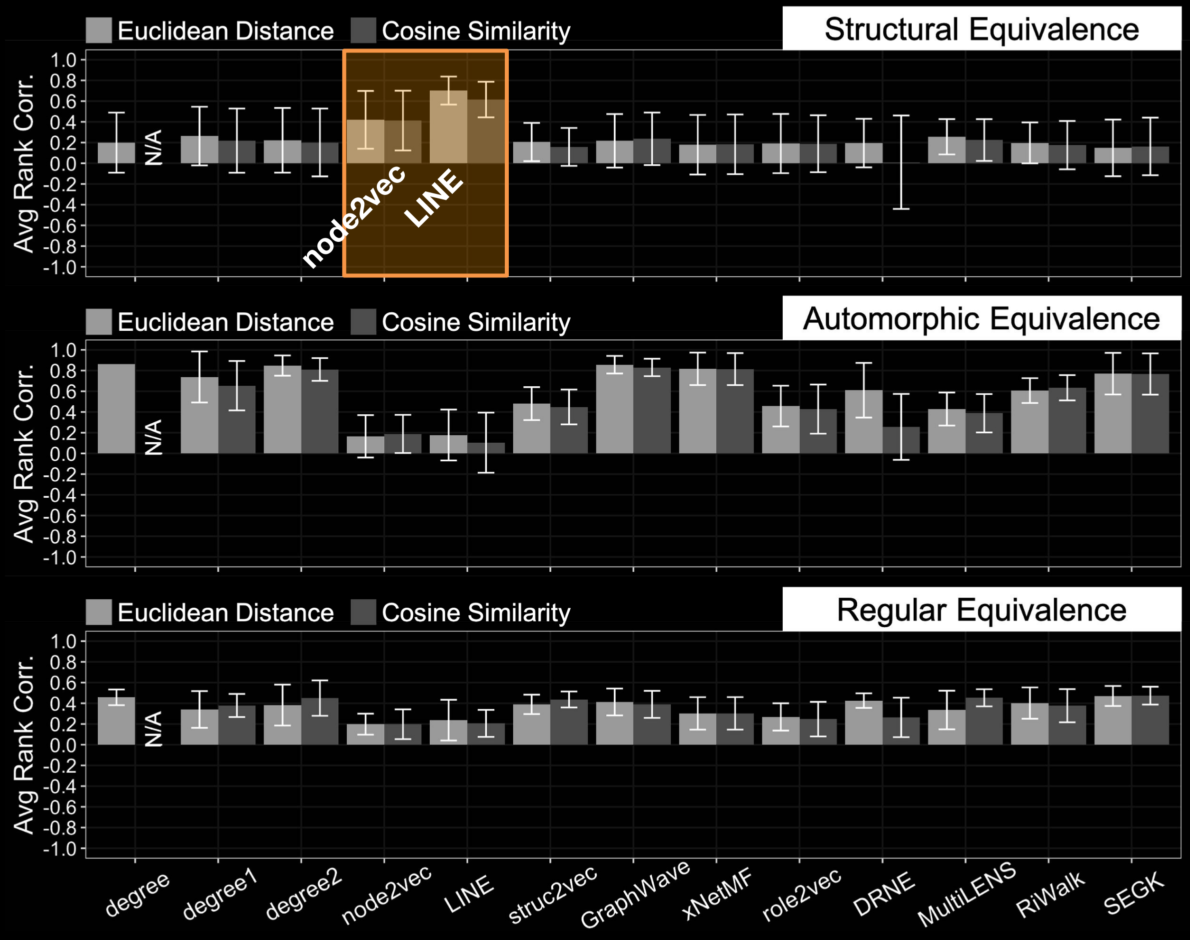
Calculate structural node properties (*intrinsic* evaluation)

Real Datasets

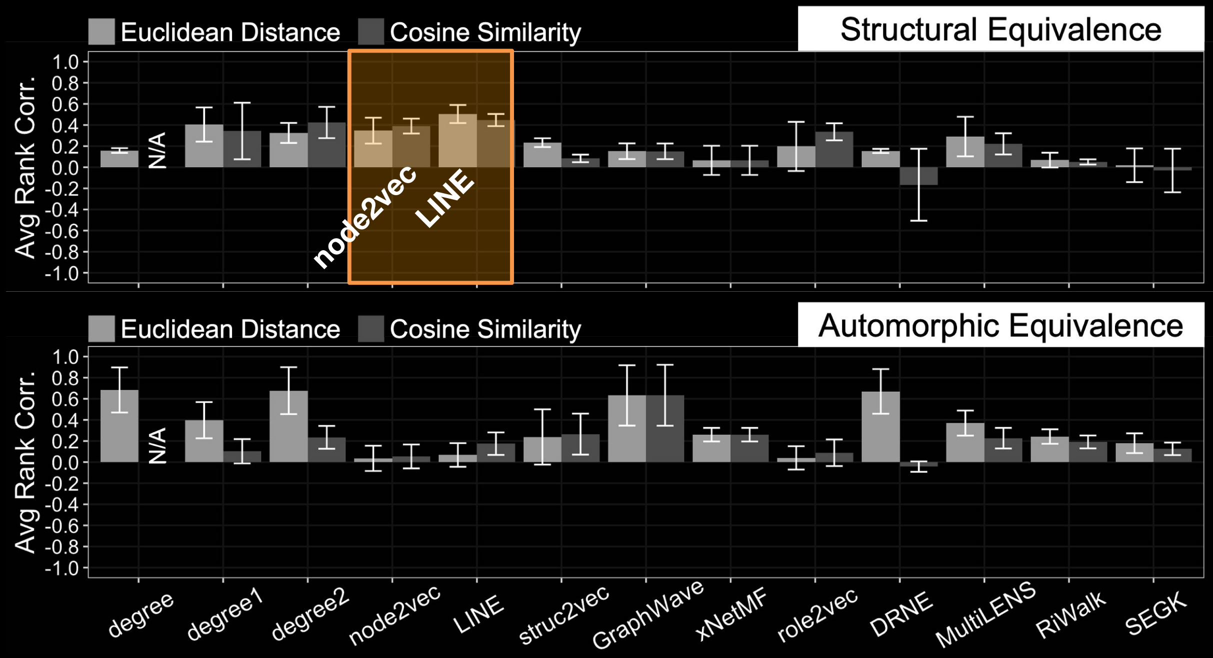
Dataset	# Nodes	# Edges	Labels
BlogCatalog	10,312	333,983	Centralities
Facebook	4,039	88,234	Equivalences

[Jin, Heimann, Jin, Koutra. 2021. Toward Understanding and Evaluating Structural Node Embeddings. ACM TKDD 2021]

Intrinsic Evaluation - Results



Synthetic Datasets

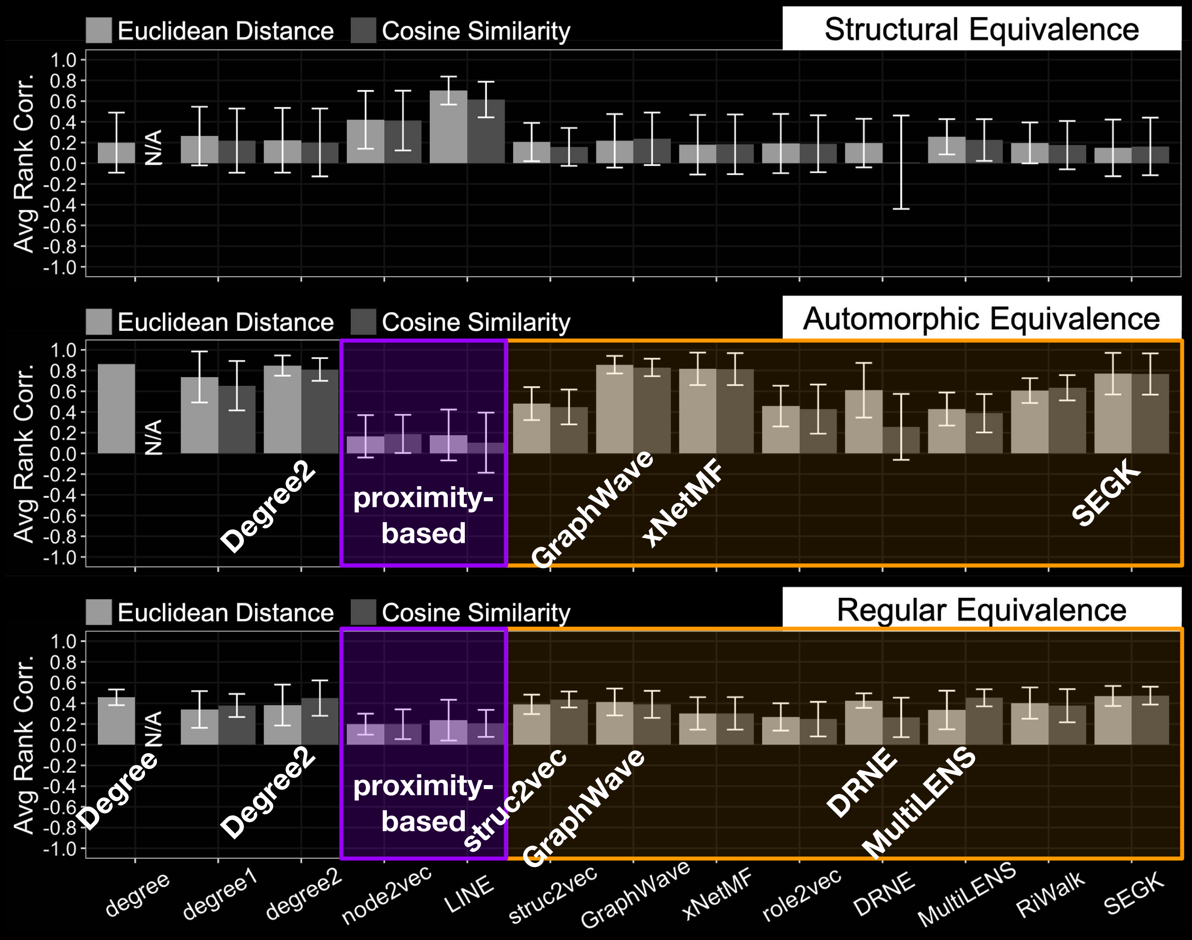


Real Datasets

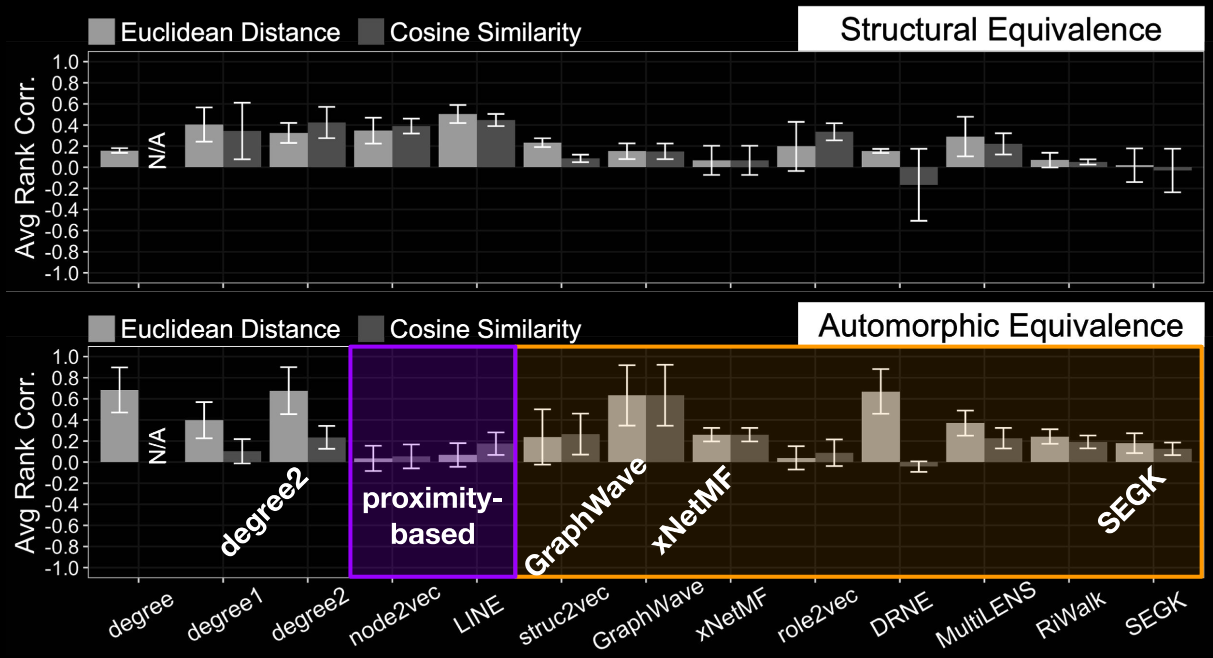
Obs. 1: LINE and node2vec rank top in structural equivalence
(as expected; based on proximity)

[Jin, Heimann, Jin, Koutra. 2021. Toward Understanding and Evaluating Structural Node Embeddings. ACM TKDD 2021]

Intrinsic Evaluation - Results



Synthetic Datasets

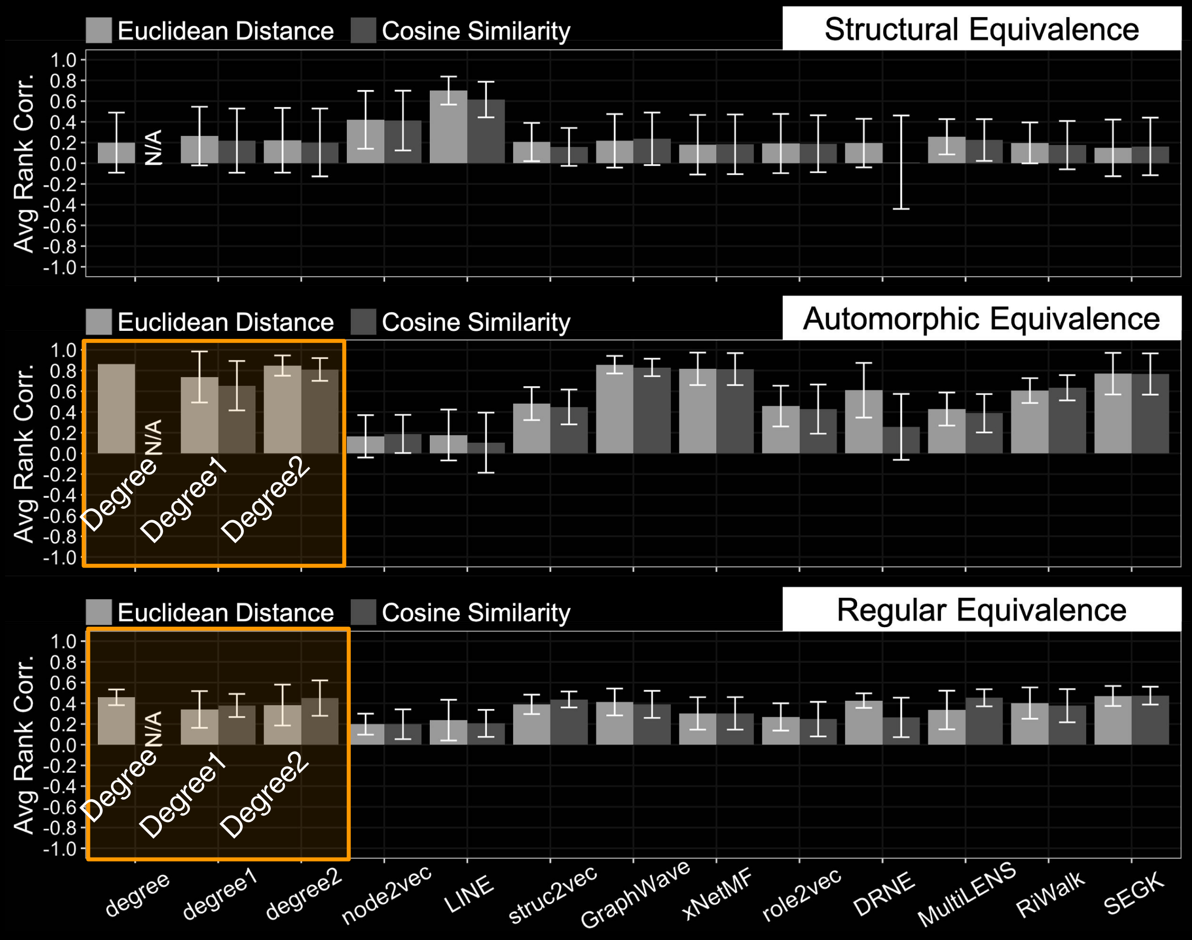


Real Datasets

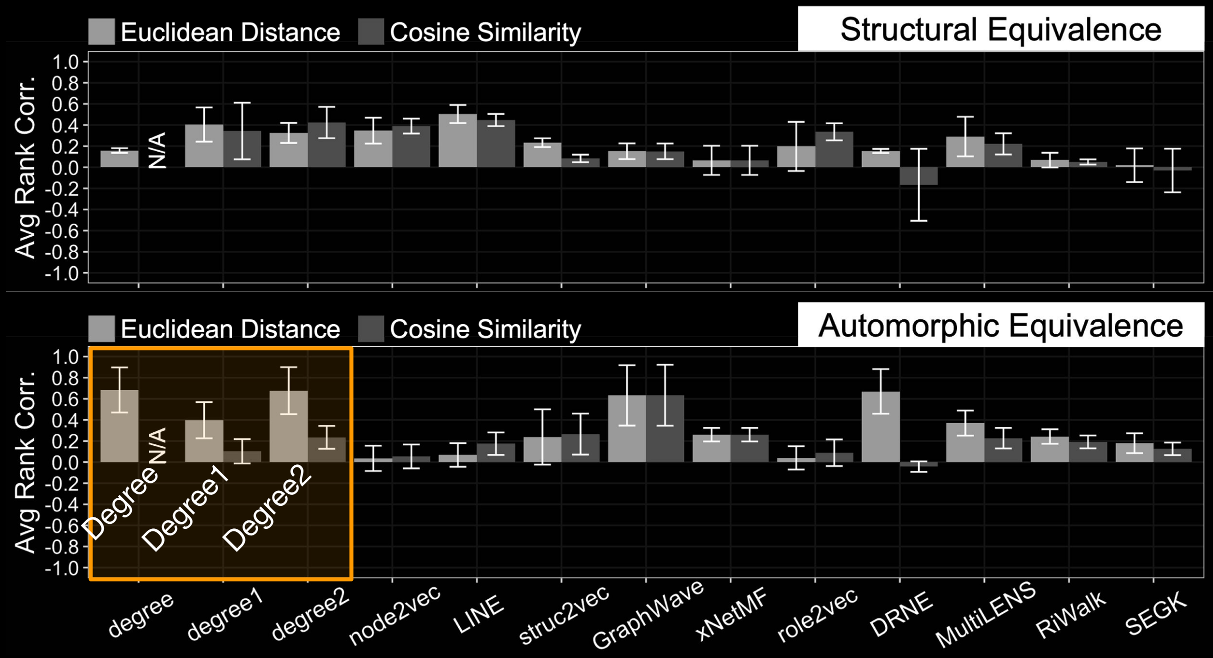
Obs. 2: Structural embedding methods do well in automorphic and regular equivalence

[Jin, Heimann, Jin, Koutra. 2021. Toward Understanding and Evaluating Structural Node Embeddings. ACM TKDD 2021]

Intrinsic Evaluation - Results



Synthetic Datasets

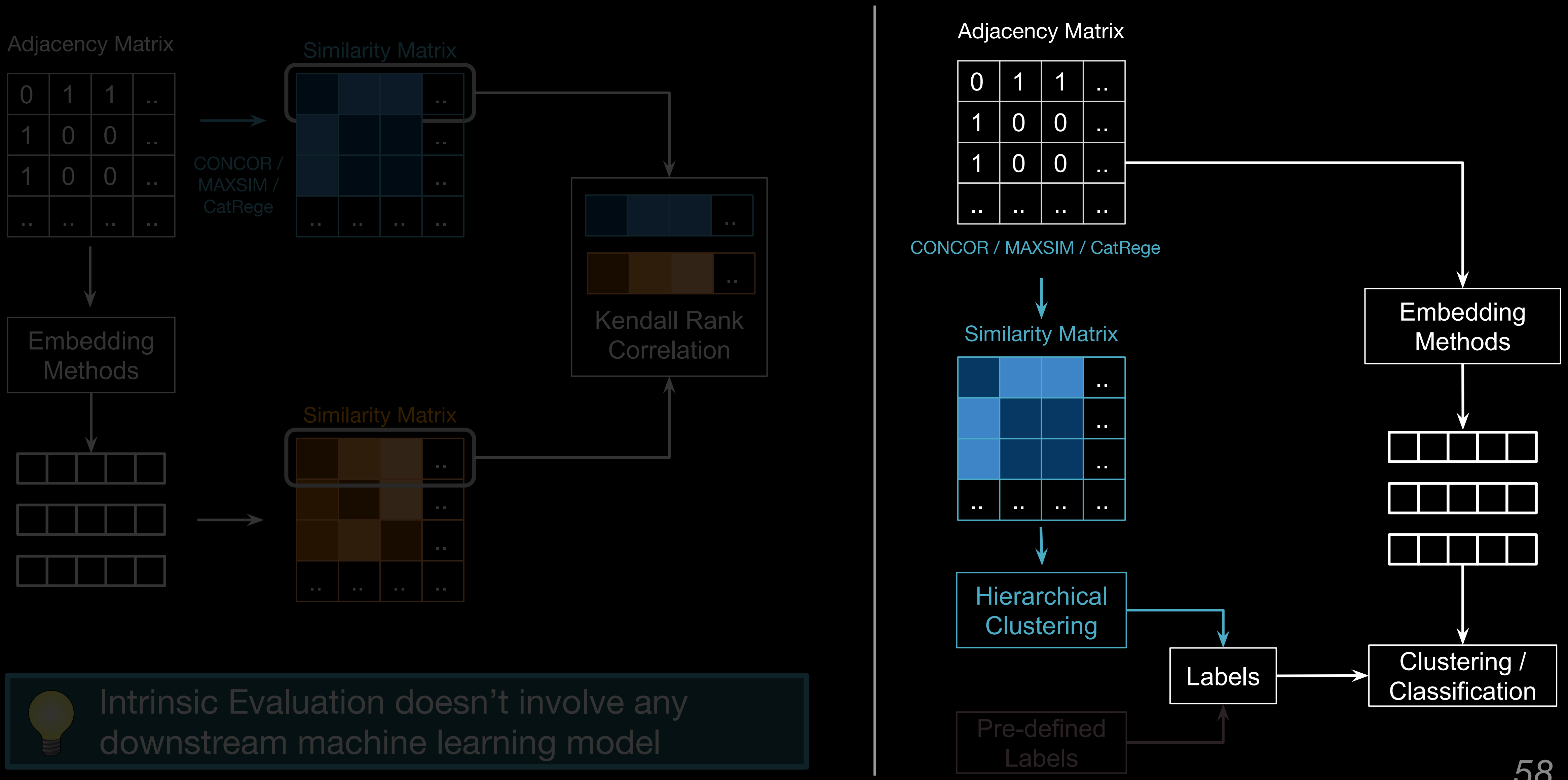


Real Datasets

Obs. 3: Degree Variants may indeed be good indicators of the structural position / role

[Jin, Heimann, Jin, Koutra. 2021. Toward Understanding and Evaluating Structural Node Embeddings. ACM TKDD 2021]

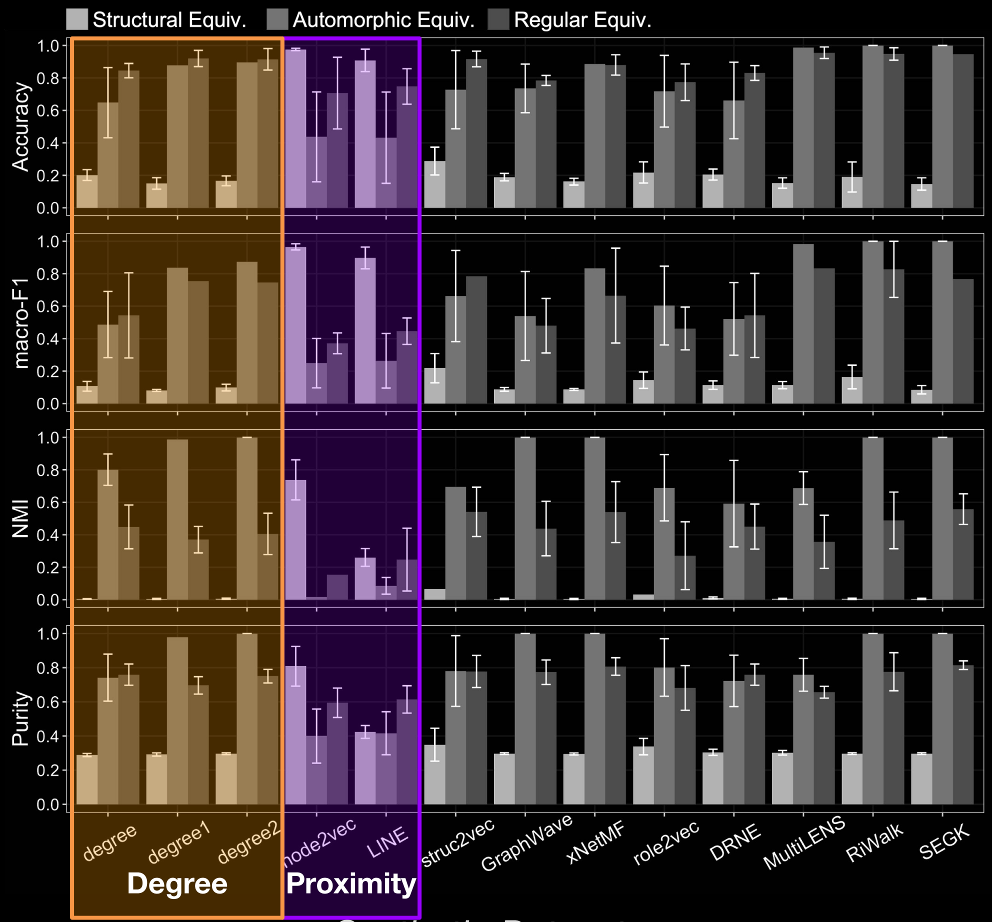
Intrinsic and Extrinsic Evaluation



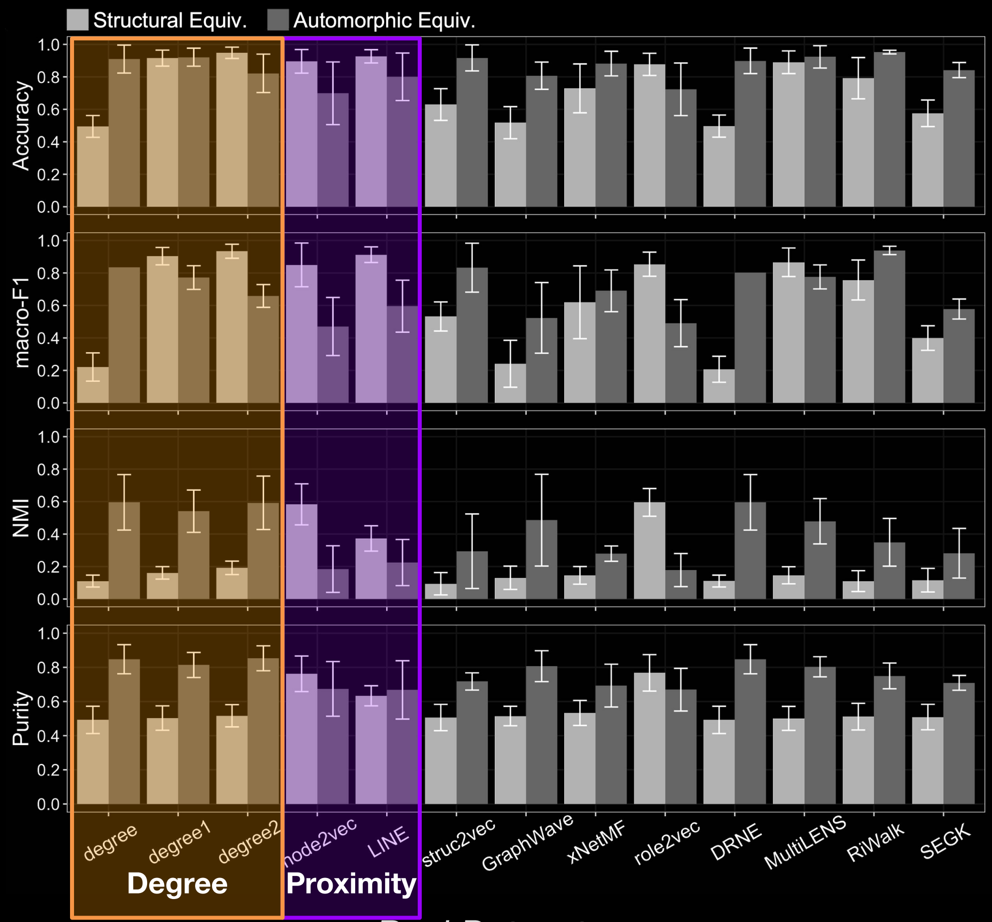
Intrinsic Evaluation doesn't involve any downstream machine learning model

[Jin, Heimann, Jin, Koutra. 2021. Toward Understanding and Evaluating Structural Node Embeddings. ACM TKDD 2021]

Extrinsic Evaluation - Results



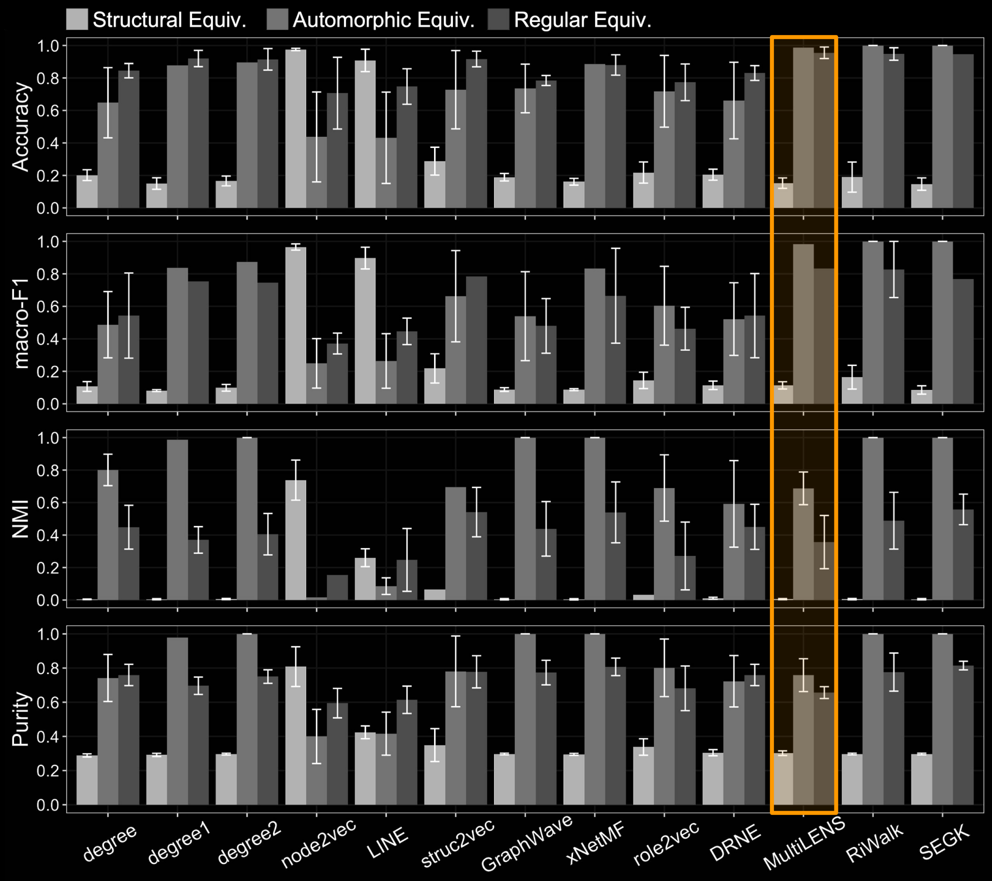
Synthetic Datasets



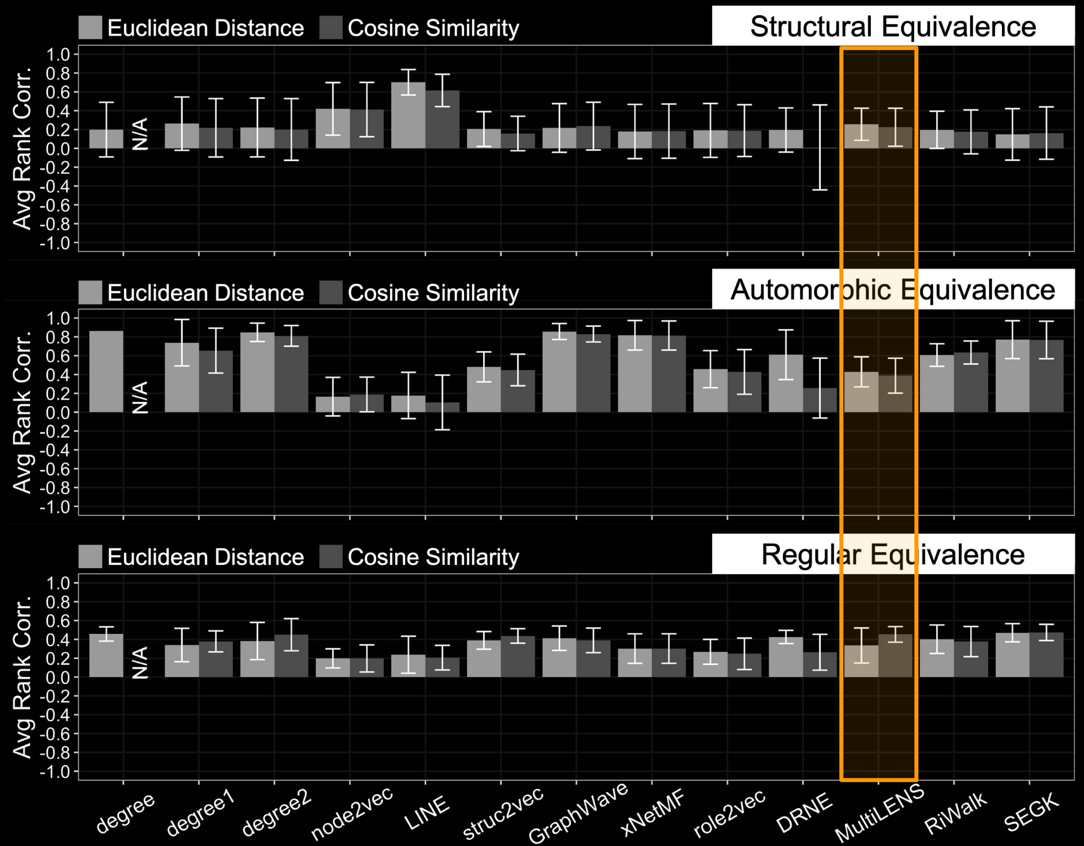
Real Datasets

Obs. 4: Similar results between intrinsic and extrinsic evaluation as well as synthetic versus real networks with one exception - MultiLENS

Extrinsic Evaluation



Synthetic Datasets - Extrinsic



Synthetic Datasets - Intrinsic

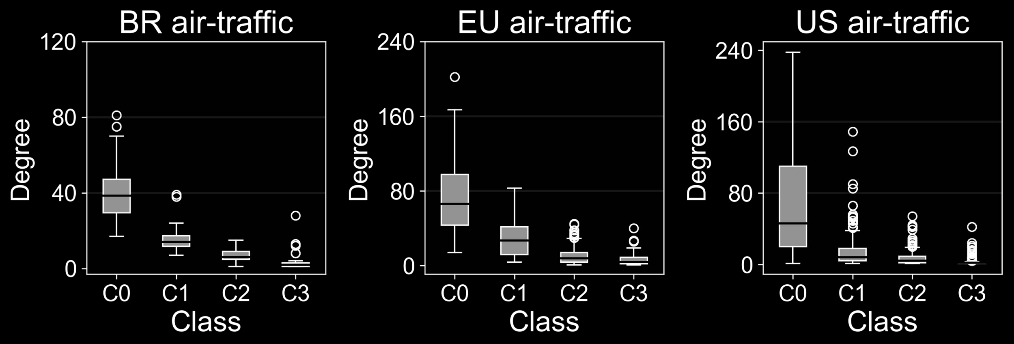
Obs. 5: Intrinsic evaluations of embeddings may not always accurately predict performance in downstream tasks - involvement of downstream ML models

[Jin, Heimann, Jin, Koutra. 2021. Toward Understanding and Evaluating Structural Node Embeddings. ACM TKDD 2021]

Issues with Label Definitions

For each airport, we assign one of four possible labels corresponding to their activity. In particular, for each dataset, we use the quartiles obtained from the empirical activity distribution to split the dataset in four groups, assigning a different label for each group. Thus, label 1 is given to the 25% less active airports, and so on. Note that all classes (labels) have the same size (number of airports). Moreover, classes are related more to the role played by the airport.

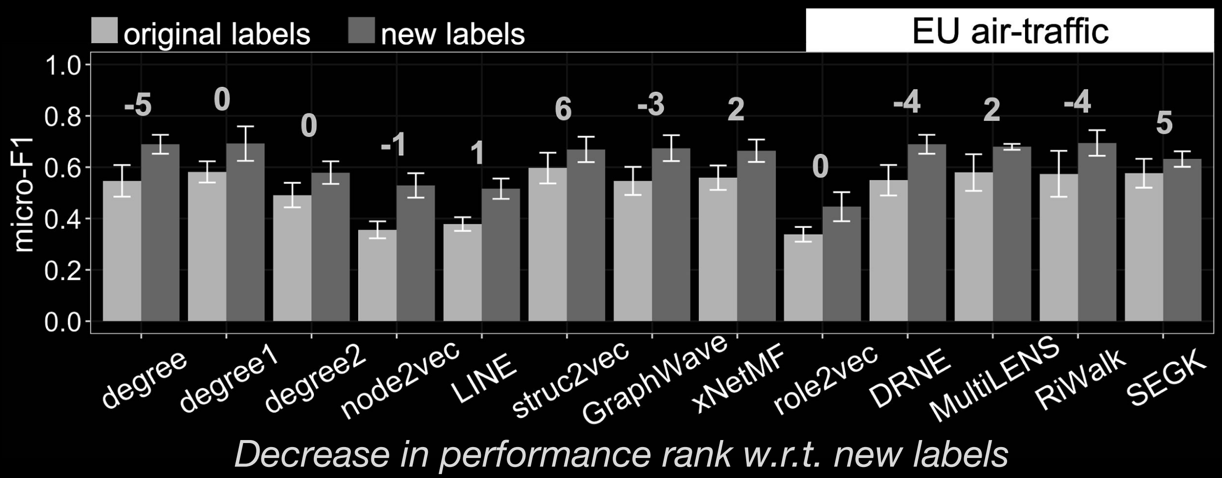
[struc2vec, Leonardo+ ‘17]



Obs. 6: Labels strongly correlated with node degree for air-traffic datasets

Original Label Splitting Manner

New Label Splitting Manner (Log → Power Law)

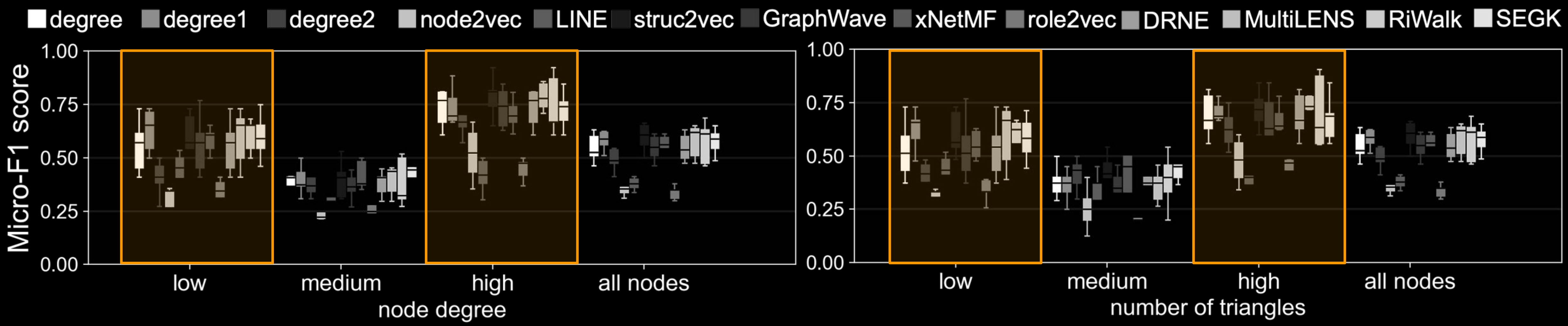


Obs. 7: Each structural embedding method best captures certain structural roles in the network, but unclear how well these roles are correlated with the labels

[Jin, Heimann, Jin, Koutra. 2021. Toward Understanding and Evaluating Structural Node Embeddings. ACM TKDD 2021]

Deeper View into Performance Scores

[0, max_property$^{1/3}$)	[max_property$^{1/3}$, max_property$^{2/3}$)	[max_property$^{2/3}$, max_property]
Clustering / Classification	Clustering / Classification	Clustering / Classification

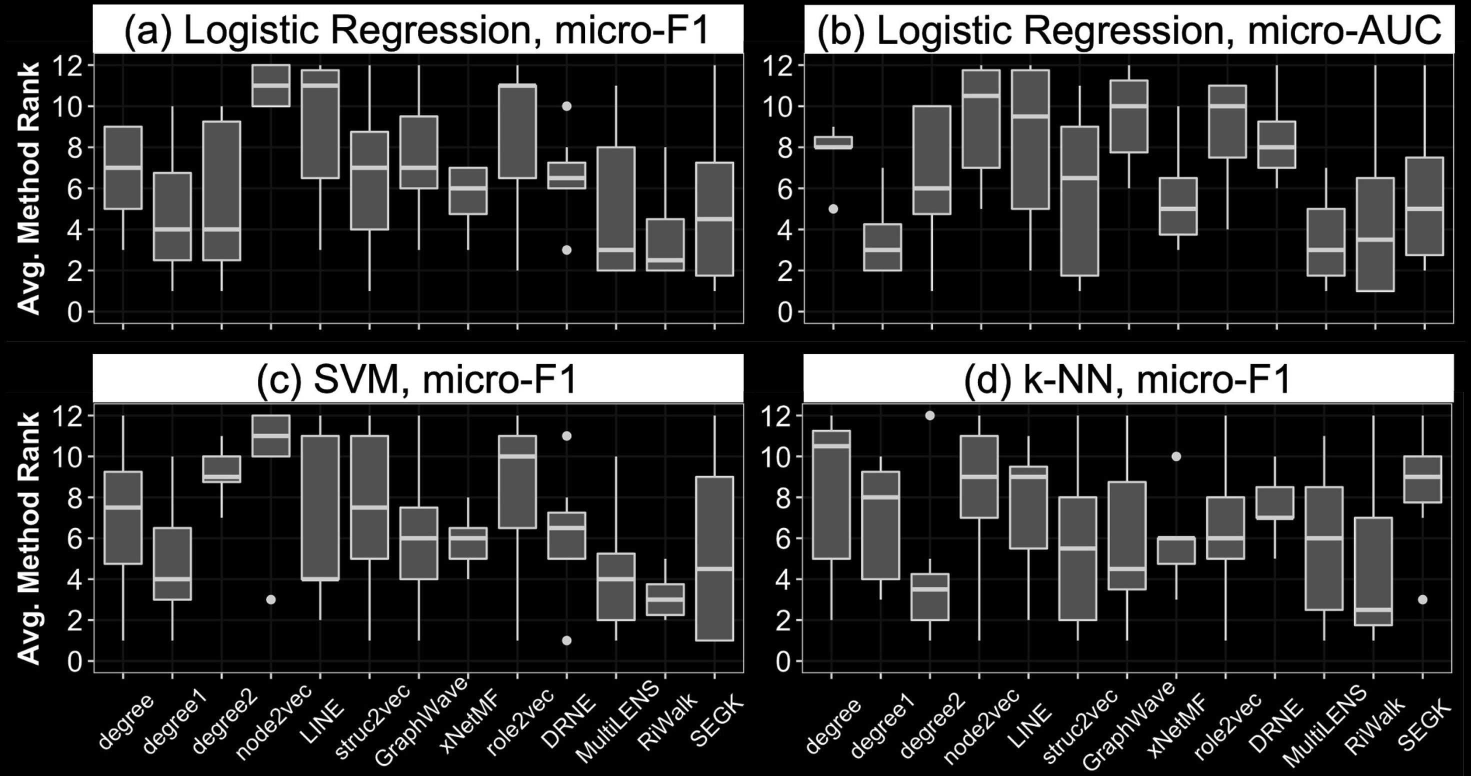


EU Air-traffic with Original Label

Obs. 8: Extreme nodes with (low/high) (degree/#triangles) tend to perform better with evaluation task

[Jin, Heimann, Jin, Koutra. 2021. Toward Understanding and Evaluating Structural Node Embeddings. ACM TKDD 2021]

Overall Performance with Pre-defined Labels



Lower is better: performance ranking summarized across all real datasets with pre-defined labels

Obs. 9: Different classifiers or performance metrics also affect performance.

Obs. 10: Methods capturing the degree distributions in local neighborhoods are among the most effective (xNetMF, MultiLENS, SEGK, degree variants)

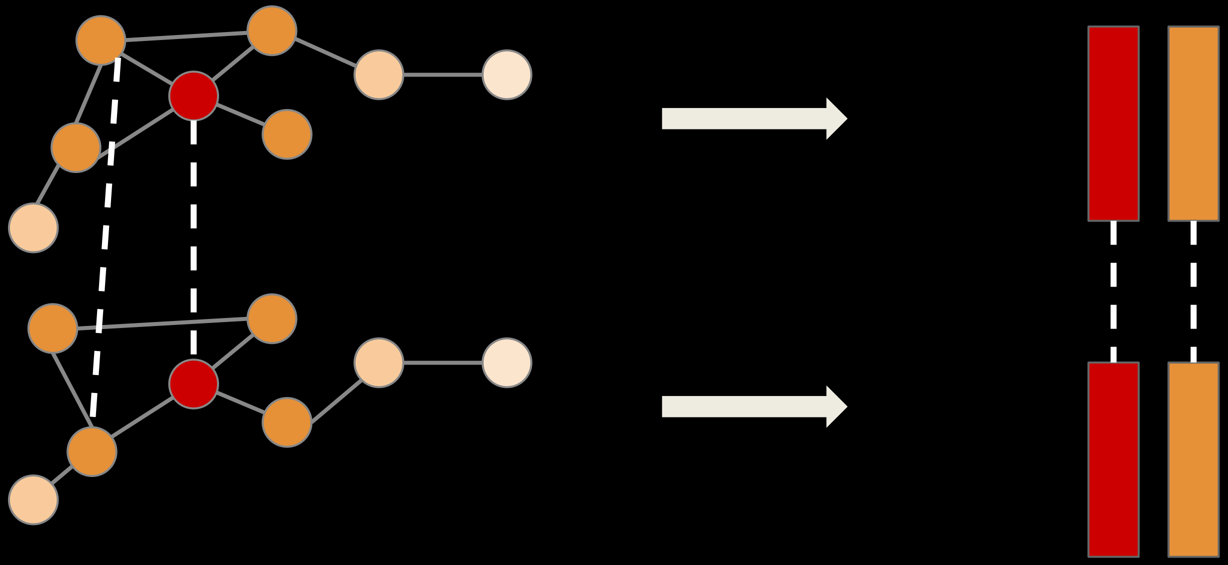
[Jin, Heimann, Jin, Koutra. 2021. Toward Understanding and Evaluating Structural Node Embeddings. ACM TKDD 2021]

Tutorial Outline:
Network Embedding for Role Discovery

- Part I: Lecture
  - Introduction
  - Structural roles in
    - network science
    - mathematical sociology
  - Structural or role-based embedding methods
  - Mining structural roles within a network
  - **Mining structural roles across networks**
- Part II: Demo
  - Hands-on demo

Embedding-based Network Alignment

Task: match corresponding nodes across networks



REGAL Framework: *Match nodes with similar structural node embeddings* [Heimann+ '18]

Observation: structural roles are often comparable across networks

[Jin, Heimann, Jin, Koutra. 2021. Toward Understanding and Evaluating Structural Node Embeddings. ACM TKDD 2021]

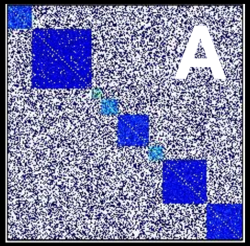
Network Alignment: Setup

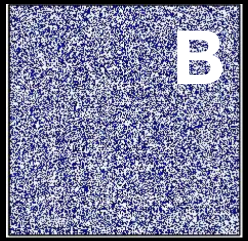
- Datasets: Networks with real-world structure from multiple domains

Name	# Nodes	# Edges	Graphs	Classes	Node labels	Domain
Arenas Email [31]	1,133	5,451	2	-	N	communication network
PPI [9]	3,890	76,584	2	-	N	PPI network (Human)

- Setup: Align graphs with adj matrices **A** and **B** = **PAP**^{T} + noise

random permutation matrix

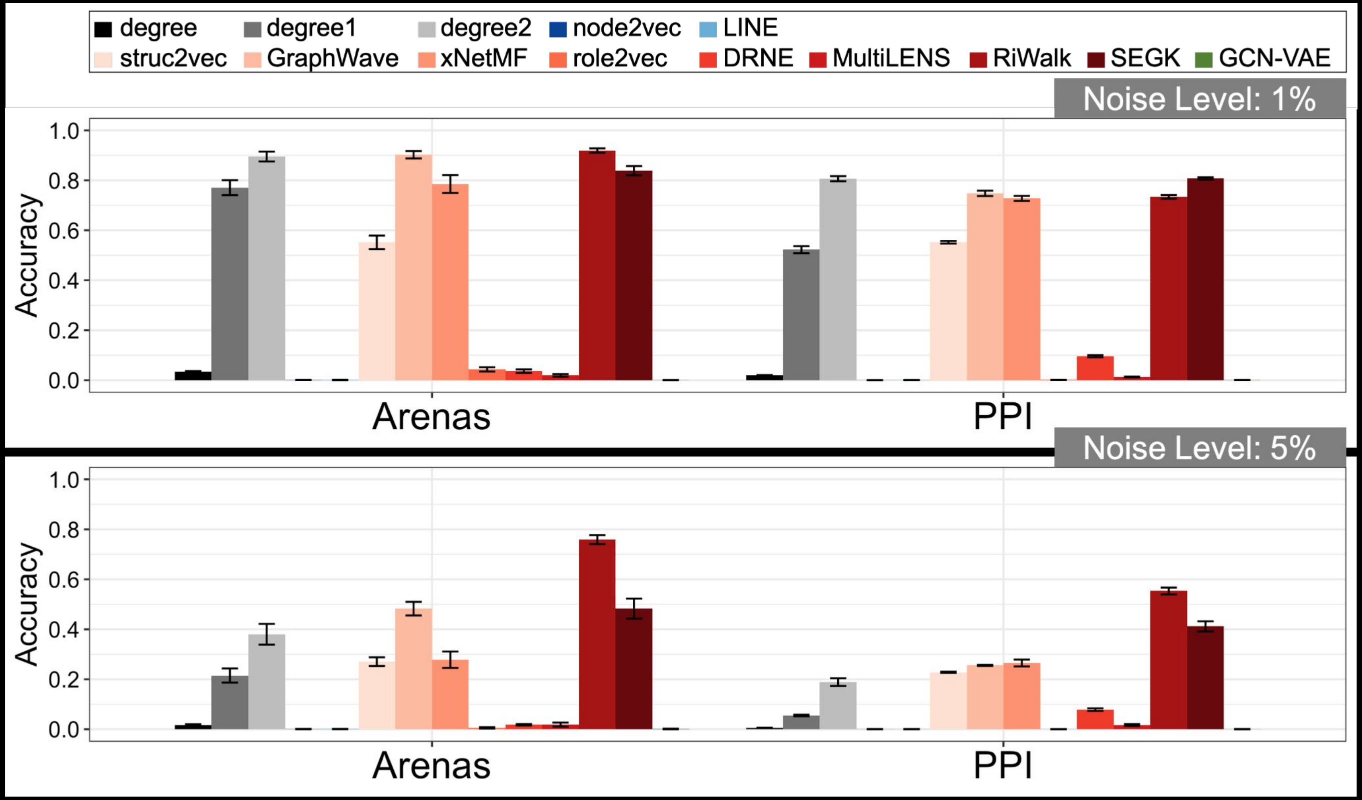




remove edges from **A** with probability p_a

[Jin, Heimann, Jin, Koutra. 2021. Toward Understanding and Evaluating Structural Node Embeddings. ACM TKDD 2021]

Network Alignment: Results



❌ Proximity-preserving methods (LINE, node2vec, GCN-VAE)

- and ones using related techniques (role2vec, DRNE, MultiLENS)

✅ xNetMF (Originally proposed), degree1 and 2

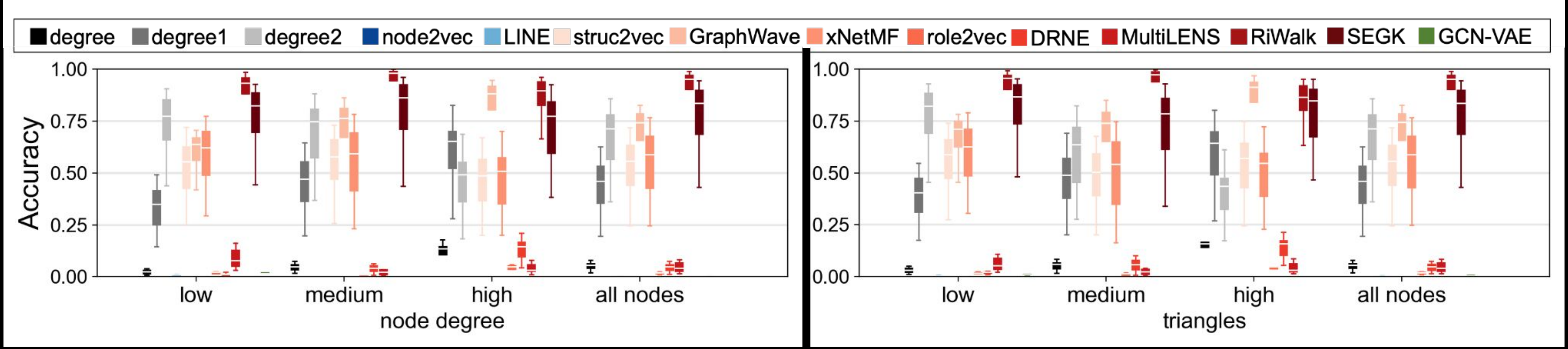
- Use higher-order local degree connectivity

✅✅ Improvement: SEGK (uses WL test to generalize notion of connectivity beyond degree)

✅✅✅ Best: RiWalk (also doesn't restrict itself to local neighborhoods)

[Jin, Heimann, Jin, Koutra. 2021. Toward Understanding and Evaluating Structural Node Embeddings. ACM TKDD 2021]

Network Alignment: Deeper Insights



Observation: Some methods (e.g. degree1, GraphWave) do better on high-connectivity nodes

- More distinguishing structural informatio\n, though also more susceptible to noise model

Observation: Best method (e.g. RiWalk) fairly consistent across connectivity level (thanks to generalizing notion of connectivity beyond degree)

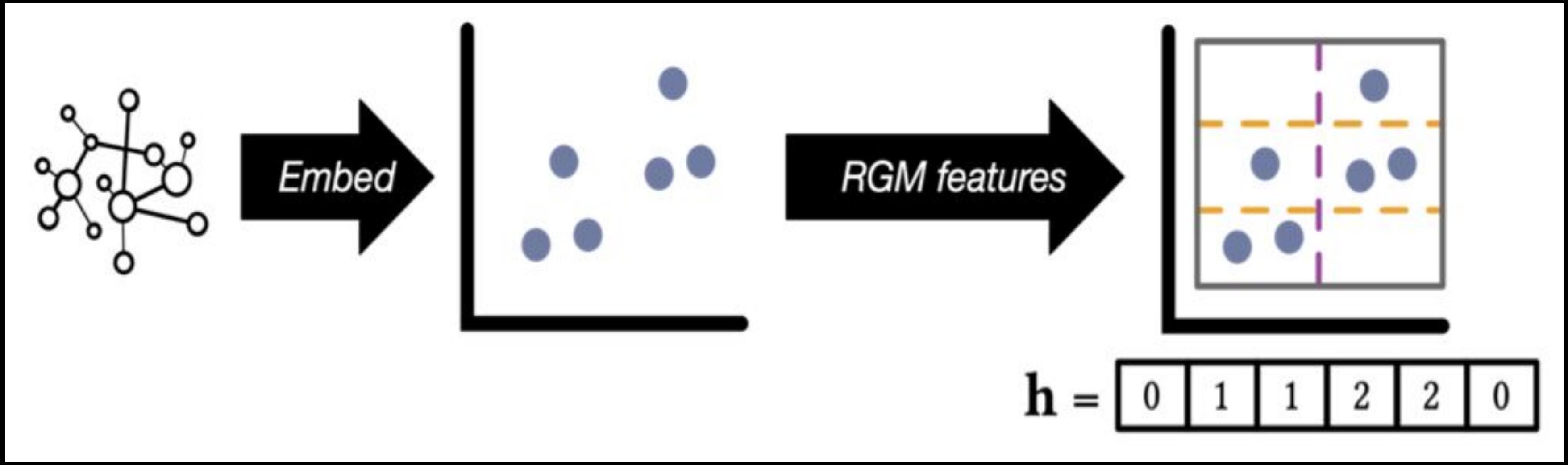
[Jin, Heimann, Jin, Koutra. 2021. Toward Understanding and Evaluating Structural Node Embeddings. ACM TKDD 2021]

Embedding-based Graph Classification

Task: predict label of entire graph (design feature vector for ML classifier)

RGM Framework: *Graph features = distribution of node features in latent space*

- Observation: structural roles are often comparable across networks [Heimann+ '19]



[Jin, Heimann, Jin, Koutra. 2021. Toward Understanding and Evaluating Structural Node Embeddings. ACM TKDD 2021]

Graph Classification: Setup

- Datasets: Common graph classification benchmarks from multiple domains

Name	# Nodes	# Edges	Graphs	Classes	Node labels	Domain
PTC-MR [37]	4,916	5,053	344	2	Y	bioinformatics
IMDB-M [37]	19,502	98,910	1,500	3	N	collaboration
NCI1 [37]	122,765	132,753	4,110	2	Y	bioinformatics

- Setup: Train kernel SVM on top of RGM features

[Jin, Heimann, Jin, Koutra. 2021. Toward Understanding and Evaluating Structural Node Embeddings. ACM TKDD 2021]

Graph Classification: Results

Method	PTC-MR	IMDB-M	NCI1	Average Rank
degree	56.3 ± 1.1	49.7 ± 0.9	77.5 ± 0.4	4.67
degree1	54.1 ± 1.0	54.0 ± 0.5	78.2 ± 0.1	5
degree2	55.5 ± 0.6	54.9 ± 0.4	80.0 ± 0.3	3.67
node2vec	50.0 ± 3.0	33.1 ± 0.6	53.5 ± 0.1	10.33
LINE	50.1 ± 3.1	33.3 ± 0.6	53.5 ± 0.1	9.33
struc2vec	50.0 ± 3.0	33.0 ± 0.6	53.5 ± 0.1	10.67
GraphWave	**58.5 ± 0.7**	47.2 ± 0.4	OOM	7.33
xNetMF	53.9 ± 0.6	55.5 ± 0.7	80.5 ± 0.4	3.33
role2vec	50.1 ± 3.1	33.5 ± 0.5	53.5 ± 0.1	9
DRNE	52.6 ± 1.7	47.9 ± 0.4	71.5 ± 0.2	7.33
MultiLENS	55.7 ± 1.3	54.9 ± 0.5	**82.1 ± 0.1**	**3**
RiWalk	50.0 ± 3.0	33.0 ± 0.6	53.5 ± 0.1	10.67
SEGK	53.3 ± 0.8	55.0 ± 0.6	OOM	7.33
GCN-VAE	50.3 ± 3.2	**74.4 ± 0.5**	OOM	7.33

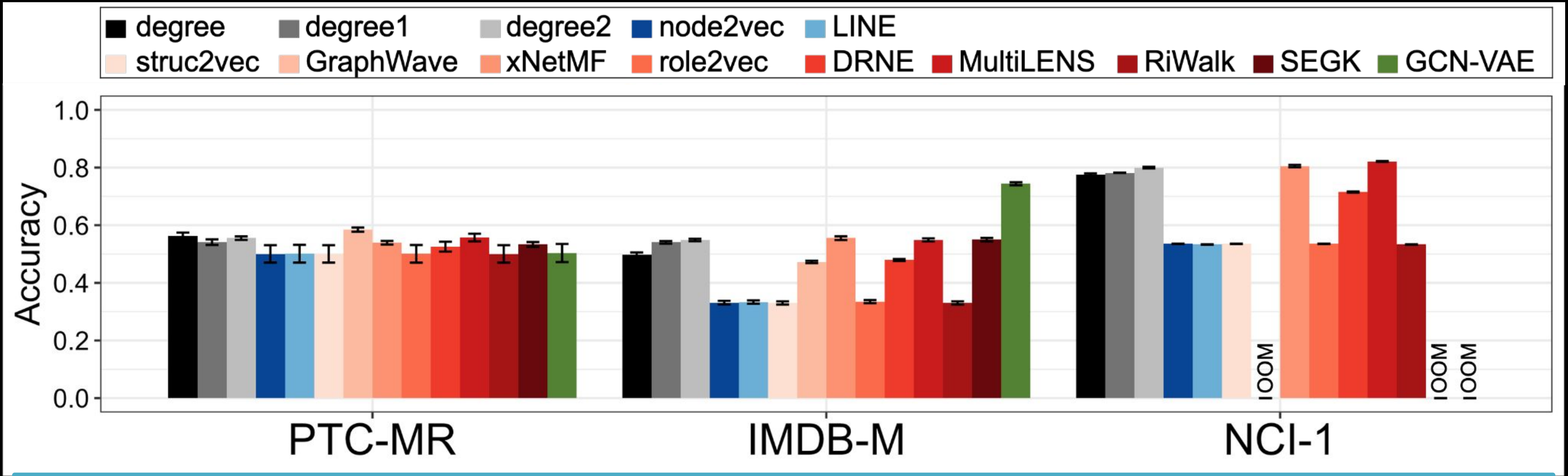
Small molecule dataset PTC-MR may have less complex structural roles, leading to similar performance for most methods

Random-walk based sampling methods perform poorly
- blur structural information too much on small graphs

Note: competitive to SOTA, e.g. GCN-VAE on IMDB-M (of independent interest)

[Jin, Heimann, Jin, Koutra. 2021. Toward Understanding and Evaluating Structural Node Embeddings. ACM TKDD 2021]

Graph Classification: Results



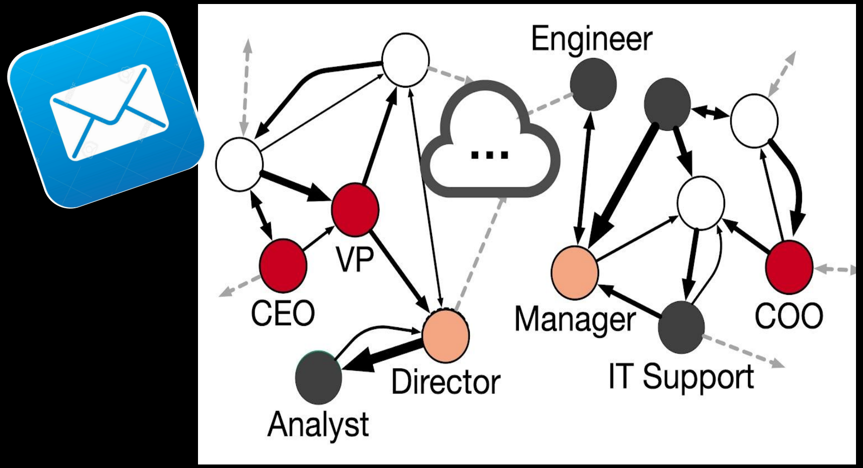
Observation: Best methods aggregate local connectivity (xNetMF, MultiLENS, SEGK, degree histogram)

Observation: Higher-order connectivity information slightly helps on larger datasets

[Jin, Heimann, Jin, Koutra. 2021. Toward Understanding and Evaluating Structural Node Embeddings. ACM TKDD 2021]

Application: Professional Role Discovery Across Companies from Email Behavior

- Hypothesis: *professional roles* of email users related to *structural roles* in email communication networks



Asymmetric communication of varying strengths

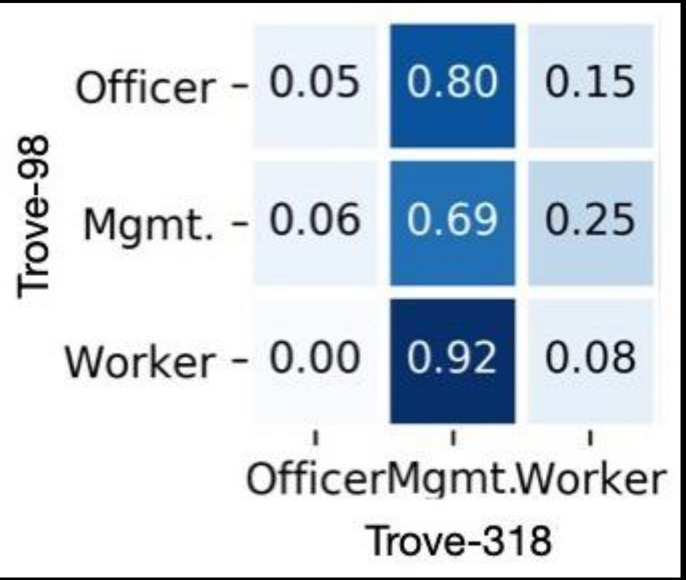
Extend xNetMF embeddings to model:

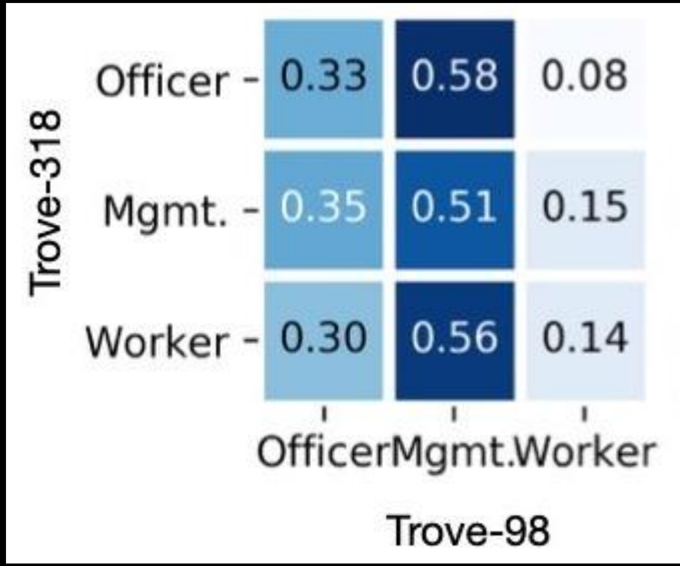
- edge weights: weigh neighbors' contributions to node's identity
- edge directions: count neighbors along incoming and outgoing edges separately

[Jin*, Heimann*, et al. KDD 2019]

Application

Comparing Roles Across Companies





Observation: Most employees at small company (Trove-98) map to lower roles at big company (Trove-318)

Observation: Most employees at big company (Trove-318) map to higher roles at small company (Trove-98)

Explanation: For a given professional rank, employees at larger companies likely more connected

[Jin*, Heimann*, et al. KDD 2019]

Application

Academic vs Corporate Hierarchy: Profs

	Officer	Mgmt.	Worker
Trove-318	0.13	0.62	0.25
Trove-183	0.24	0.51	0.26
Trove-141	0.17	0.69	0.15
Trove-98	0.57	0.31	0.11
Trove-19	0.75	0.11	0.13

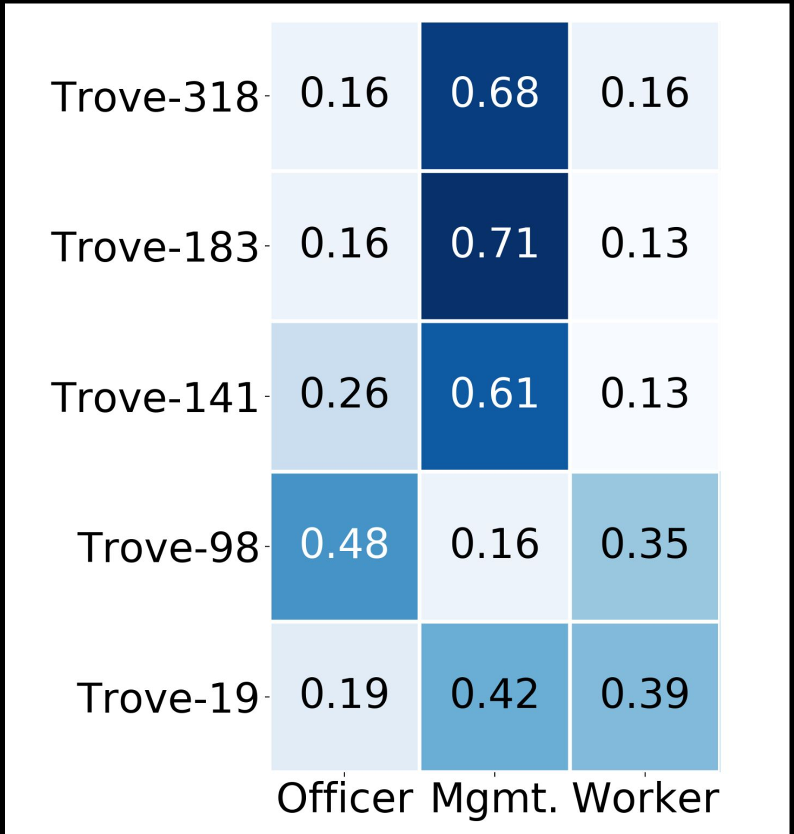
Mapping Professors to Professional Roles

Observation: Professors behave like executives of small companies / managers of large ones

[Jin*, Heimann*, et al. KDD 2019]

Academic vs Corporate Hierarchy: Students

Application



Mapping Grad Students to Professional Roles

Observation: Graduate students behave like managers/employees of other roles

[Jin*, Heimann*, Safavi+, KDD 2019]

Tutorial Outline:
Network Embedding for Role Discovery

- Part I: Lecture
  - Introduction
  - Structural roles in
    - network science
    - mathematical sociology
  - Structural or role-based embedding methods
  - Mining structural roles within a network
  - Mining structural roles across networks
- Part II: Demo
  - Hands-on demo

Part I: Take-away messages

- Structural / role-based embeddings and equivalence types from sociology
- Intrinsic and extrinsic evaluation of embeddings
  - *Structural* equivalence best captured by proximity-based methods
  - Structural embedding methods better capture *automorphic* and *regular* equivalence
  - Degree variants can be building blocks for future methods
- Comparison of structural embeddings for single and multi-network analysis

Tutorial Outline:
Network Embedding for Role Discovery

- Part I: Lecture
  - Introduction
  - Structural roles in
    - network science
    - mathematical sociology
  - Structural or role-based embedding methods
  - Mining structural roles within a network
  - Mining structural roles across networks
- **Part II: Demo**
  - **Hands-on demo**

Hands On: Structural Embeddings Graph Library

https://github.com/GemsLab/StrucEmbedding-GraphLibrary

README.md

The Structural EMBedding graph library (SEMB)

Authors: GEMS Lab Team @ University of Michigan (Mark Jin, Ruowang Zhang, Mark Heimann)

This SEMB library allows fast onboarding to get and evaluate structural node embeddings. With the unified API interface and the modular codebase, SEMB library enables easy intergration of 3rd-party methods and datasets.

The library itself has already included a set of popular methods and datasets ready for immediate use.

- Built-in methods: node2vec, struc2vec, GraphWave, xNetMF, role2vec, DRNE, MultiLENS, RiWalk, SEGK, (more methods to add in the near future)
- Built-in datasets:

Dataset	# Nodes	# Edges
BlogCatalog	10,312	333,983
Facebook	4,039	88,234
ICEWS	1,255	1,414
PPI	56,944	818,786
BR air-traffic	131	1,038
EU air-traffic	399	5,995
US air-traffic	1,190	13,599
DD6	4,152	20,640
Synthetic Datasets		

References: Structural Role-based Embeddings

- Leonardo FR Ribeiro, Pedro HP Saverese, and Daniel R Figueiredo. struc2vec: Learning node representations from structural identity. In KDD, 2017.
- Nesreen K. Ahmed, Ryan A. Rossi, John Boaz Lee, Theodore L. Willke, Rong Zhou, Xiangnan Kong, and Hoda Eldardiry. role2vec: Role-based network embeddings. In DLG KDD, 2019.
- Claire Donnat, Marinka Zitnik, David Hallac, and Jure Leskovec. Learning structural node embeddings via diffusion wavelets. In KDD, 2018.
- Aditya Grover and Jure Leskovec. node2vec: Scalable feature learning for networks. In KDD. ACM, 2016.
- Mark Heimann, Haoming Shen, Tara Safavi, and Danai Koutra. Regal: Representation learning-based graph alignment. In 2018 ACM International Conference on Information and Knowledge Management (CIKM). ACM, 2018.
- Di Jin, Ryan A Rossi, Eunyee Koh, Sungchul Kim, Anup Rao, and Danai Koutra. Latent network summarization: Bridging network embedding and summarization. In KDD, 2019.
- Junchen Jin, Mark Heimann, Di Jin, and Danai Koutra. Towards understanding and evaluating structural node embeddings. In ACM Trans. Knowl. Discov. Data. ACM, 2021.
- Giannis Nikolentzos and Michalis Vazirgiannis. Learning structural node representations using graph kernels. TKDE, 2019.
- Ma Xuewei, Geng Qin, Zhiyang Qiu, Mingxin Zheng, and Zhe Wang. Riwalk: Fast structural node embedding via role identification. IEEE ICDM 2019.

References: Structural Role-based Embeddings

- Jian Tang, Meng Qu, Mingzhe Wang, Ming Zhang, Jun Yan, and Qiaozhu Mei. LINE: Large-scale information network embedding. In WWW, 2015.
- Aditya Grover and Jure Leskovec. node2vec: Scalable feature learning for networks. In KDD, 2016.
- Thomas N Kipf and Max Welling. Variational graph auto-encoders. NIPS Workshop on Bayesian Deep Learning, 2016.
- Ke Tu, Peng Cui, Xiao Wang, Philip S. Yu, and Wenwu Zhu. Deep recursive network embedding with regular equivalence. In KDD, pages 2357–2366, 2018.
- Di Jin, Mark Heimann, Tara Safavi, Mengdi Wang, Wei Lee, Lindsay Snider, and Danai Koutra. Smart roles: Inferring professional roles in email networks. In KDD, 2019.
- Mark Heimann, Tara Safavi, and Danai Koutra. Distribution of node embeddings as multiresolution features for graphs. In 2019 IEEE International Conference on Data Mining (ICDM). IEEE, 2019.
- Ryan A. Rossi, Di Jin, Sungchul Kim, Nesreen K. Ahmed, Danai Koutra, and John Boaz Lee. On proximity and structural role-based embeddings in networks: Misconceptions, techniques, and applications. ACM Trans. Knowl. Discov. Data, 14(5):63:1–63:37, 2020.
- Jing Zhu, Xingyu Lu, Mark Heimann, and Danai Koutra. Node proximity is all you need: Unified structural and positional node and graph embedding. In SDM, 2021.

References: Mathematical Sociology

- S.P. Borgatti and M.G. Everett. 1992. Notions of position in social network analysis. Sociological methodology22, 1 (1992)
- Stephen P Borgatti, Martin G Everett, and Jeffrey C Johnson. 2018. Analyzing social networks. Sage
- F. Lorrain and H.C. White. 1971. Structural equivalence of individuals in social networks. Journal of Mathematical Sociology
- S. Boorman, H.C. White: Social Structure from Multiple Networks: II. Role Structures. American Journal of Sociology, 81:1384-1446, 1976.
- R.S. Burt: Positions in Networks. Social Forces, 55:93-122, 1976.
- M.G. Everett, S. P. Borgatti: Regular Equivalence: General Theory. Journal of Mathematical Sociology, 19(1):29-52, 1994.
- K. Faust, A.K. Romney: Does Structure Find Structure? A critique of Burt's Use of Distance as a Measure of Structural Equivalence. Social Networks, 7:77-103, 1985.
- K. Faust, S. Wasserman: Blockmodels: Interpretation and Evaluation. Social Networks, 14:5–61. 1992.
- R.A. Hanneman, M. Riddle: Introduction to Social Network Methods. University of California, Riverside, 2005.
- L.D. Sailer: Structural Equivalence: Meaning and Definition, Computation, and Applications. Social Networks, 1:73-90, 1978.
- M.K. Sparrow: A Linear Algorithm for Computing Automorphic Equivalence Classes: The Numerical Signatures Approach. Social Networks, 15:151-170, 1993.
- S. Wasserman, K. Faust: Social Network Analysis: Methods and Applications. Cambridge University Press, 1994.
- H.C. White, S. A. Boorman, R. L. Breiger: Social Structure from Multiple Networks I. Blockmodels of Roles and Positions. American Journal of Sociology, 81:730-780, 1976.
- D.R. White, K. Reitz: Graph and Semi-Group Homomorphism on Networks and Relations. Social Networks, 5:143-234, 1983.
- …